The detection of meningococcal disease through identification of antimicrobial peptides using an *in silico* model creation.

A thesis submitted in fulfilment of the requirement for the degree of

Philosophiae Doctor

In the Department of Biotechnology

Faculty of Natural Sciences, University of the Western Cape

Gadija Abdullah

Supervisor: Dr Ashley Pretorius

Co-supervisor: Dr Riaan Den Haan

May 2019

DECLARATION

I declare that this thesis is a presentation of my original research work. I have written the enclosed PhD Thesis, '*The detection of meningococcal disease through identification of antimicrobial peptides using an in silico model creation'* completely by myself and have documented all sources and material used. This thesis was not previously presented to another examination board in any other university and has not been published.

Gadija Abdullah May 2019

……

Signature

UNIVERSITY of the
WESTERN CAPE

ACKNOWLEGDEMENTS

To ALLAH (swt), the Cherisher and Sustainer of the Worlds for His Infinite Blessings and Guidance in my life and that of my family, in whom I trust, for all these doors did not open by accident. Shukran (Thank You).

I sincerely thank Associate Professor A. Pretorius, Chairperson of the Science Faculty, Assessment Committee and Principle Investigator of the Bioinformatics Research Group on Disease Diagnostics at the University of the Western Cape, South Africa. I am grateful and appreciative to him for the constant assistance, guidance and help in developing my research skills.

A special Shukran goes to Shuaib Abdullah, a loving husband for his support. I would not have survived these years of study without his constant encouragement and unconditional love. He has taught me that I could be anything in spite of whatever circumstances in which I find myself. He has in every sense earned this honour more than me. And to my wonderful kids, Yaghya, Yusuf, Noorah and baby Nuhaa, who throughout my studies somehow understood and complained far less. To my mom, thank you for all the support with the kids.

I thank the University of the Western Cape for this opportunity to further my studies.

I would like to thank my fellow students at the Bioinformatics Research Group (BRG) of the University of the Western Cape for their insights, suggestions and training on various databases which assisted me in this research study.

ABSTRACT

The detection of meningococcal disease through identification of antimicrobial peptides using an *in silico* model creation.

G. Abdullah

Ph.D. thesis, Department of Biotechnology, Faculty of Natural Sciences, University of the Western Cape

Neisseria meningitidis (the meningococcus), the causative agent of meningococcal disease (MD) was identified in 1887 and despite effective antibiotics and partially effective vaccines, *Neisseria meningitidis* (*N. meningitidis*) is the leading cause worldwide of meningitis and rapidly fatal sepsis usually in otherwise healthy individuals. Over 500 000 meningococcal cases occur every year. These numbers have made bacterial meningitis a top ten infectious cause of death worldwide. MD primarily affects children under 5 years of age, although in epidemic outbreaks there is a shift in disease to older children, adolescents and adults. MD is also associated with marked morbidity including limb loss, hearing loss, cognitive dysfunction, visual impairment, educational difficulties, developmental delays, motor nerve deficits, seizure disorders and behavioural problems. Antimicrobial peptides (AMPs) are molecules that provide protection against environmental pathogens, acting against a large number of microorganisms, including bacteria, fungi, yeast and virus. AMPs production is a major component of innate immunity against infection. The chemical properties of AMPs allow them to insert into the anionic cell wall and phospholipid membranes of microorganisms or bind to the bacteria making it easily detectable for diagnostic purposes. AMPs can be exploited for the generation of novel antibiotics, as biomarkers in the diagnosis of inflammatory conditions, for the manipulation of the inflammatory process, wound

healing, autoimmunity and in the combat of tumour cells. Due to the severity of meningitis, early detection and identification of the strain of *N. meningitidis* is vital. Rapid and accurate diagnosis is essential for optimal management of patients and a major problem for MD is its diagnostic difficulties and experts conclude that with an early intervention the patient' prognosis will be much improved. It is becoming increasingly difficult to confirm the diagnosis of meningococcal infection by conventional methods. Although polymerase chain reaction (PCR) has the potential advantage of providing more rapid confirmation of the presence of the bacterium than culturing, it is still time consuming as well as costly. Introduction of AMPs to bind to *N. meningitidis* receptors could provide a less costly and time consuming solution to the current diagnostic problems. World Health Organization (WHO) meningococcal meningitis program activities encourage laboratory strengthening to ensure prompt and accurate diagnosis to rapidly confirm the presence of MD.

This study aimed to identify a list of putative AMPs showing antibacterial activity to *N. meningitidis* to be used as ligands against receptors uniquely expressed by the bacterium and for the identified AMPs to be used in a Lateral Flow Device (LFD) for the rapid and accurate diagnosis of MD.

Various computational AMP databases such as APD, CAMP, DRAMP and DBAASP were explored to identify a list of experimentally validated AMPs against *N. meningitidis*. The identified peptides were used to construct probabilistic models; using an *in silico* mathematical algorithm Hidden Markov Models (HMM) to be used in the scanning of various genome sequences to identify putative AMPs for *N. meningitidis.* The predicted AMPs were selected based on their E-values and having a single domain. In this study, nine AMPs (YYNN1 – YYNN9) were identified as possible anti-*N. meningitidis* peptides displaying E-values < 0.01, with the smallest E-values seen for YYNN8 and then YYNN9.

Various search engines were accessed to identify receptors in the outer membrane of *N. meningitidis* which will serve as targets to the putative AMPs. Three *N. meningitidis* proteins were shortlisted as receptors for the identified peptides namely NhhA, Opc and PorA. The major considerations for using the aforementioned proteins was based on their unique expression in *N. meningitidis* as to ensure selectivity of the bacterium during diagnosis as well the nature of expression of these proteins being extracellular, to ensure their availability for binding during diagnosis.

The physicochemical properties of the identified peptides were determined using APD and Bactibase to determine whether these peptides conformed to known AMPs since these peptides were considered putative/novel.

Furthermore, the 3D structures of the AMPs and *N. meningitidis* receptors were modelled using I-TASSER. The 3D modelling results revealed that, all the AMPs as well as the receptors were of good models structurally based on their TM, RMSD and C-score values. The AMPs showed secondary structures including α-helices and extended shapes similar to other known AMPs whereas the receptors showed structures as seen within the literature.

The 3D structures of the AMPs were docked against the 3D structures of the *N. meningitidis* receptors using PatchDock, to determine their binding affinity as well as binding orientation. The results showed that AMP YYNN5 has the highest binding affinity score of 15396 when bound to PorA as well as the highest binding affinity score when docked to NhhA with a score of 11904, whilst AMP YYNN8 showed the highest binding affinity score when docked to Opc, with a binding score of 11546 as calculated by PatchDock.

Lastly, identification of mutation sensitive or "hotspot" amino acid residues that are involved in forming the interface between the AMPs and the receptor proteins were identified using KFC. Non "hotspot" residues were changed using site directed mutagenesis with the parental AMPs as templates to generate derivative AMPs that displayed increased predicted binding affinity for the NhhA, Opc and PorA proteins. Derivative AMP YYNN2c had an increase in binding affinity from 14446 to 15072 with a net positive percentage of 4.3% when bound to the PorA protein, resulting in this AMP as having the second highest binding affinity score after YYNN5 (15396) bound to PorA although the binding affinity of YYNN5 did not increase following site directed mutagenesis.

To ensure that site directed mutagenesis did not alter the physicochemical properties or the structure of the derived AMPs, APD and Bactibase were used once more as well as I-TASSER. From the results obtained, the derived AMPs still conformed to known AMPs as well as displaying similar α-helical secondary structures as their parental counterparts, with slight variation in partial α-helical structure for certain derived AMPs.

Taken together the results of this work indicates that YYNN5 bound to the *N. meningitidis* receptor PorA is the most likely AMP to be used in a LFD for the accurate and sensitive diagnosis of MD within patient samples such as blood.

This is the first report on the identification of AMPs, using *in silico* analysis, as ligands to the bacteria *N. meningitidis* receptor proteins and can be a key step in the creation of an effective and universal diagnostic Point-of-Care (POC) method to diagnose MD at an early stage to prevent serious complications.

Keywords: Antimicrobial Peptides, *Neisseria meningitidis*, Antimicrobial Peptide Database, Hidden Markov Model, Meningococcal disease, diagnostic, database.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2D	:	2 dimensional
3D	:	3 dimensional
ABCs	:	Active Bacterial Core surveillance
ACP	:	Adhesin Complex Protein
AMPs	:	Antimicrobial peptides
APD	:	Antimicrobial Peptide Database
APIs	:	Application Programming Interfaces
App	:	Adhesion and Penetration Protein
AT	:	Autotransporter
AuNPs	:	Gold nanoparticles
AusI	:	Autotransporter serine protease I
b2AR	:	b2-adrenergic receptor
BAM	:	β-Barrel Assembly Machinery
BBB	:	Blood Brain Barrier
BD	:	Biological databases
BLAST	:	Basic Local Alignment Search Tool
CAMP	:	Collection of Anti-Microbial Peptides
CASP	:	Critical Assessment of Structure Prediction
CD66e	:	Carcinoembryonic antigen
CEACAMs	:	Carcinoembryonic antigen-related cell adhesion molecules
CNS	:	Central Nervous System
CRDM	:	Centre for Respiratory Diseases and Meningitis
CRP	:	C-reactive protein

C-score	:	Confidence score
CSF	:	Cerebral Spinal Fluid
DBAASP	:	Database of Antimicrobial Activity and Structure of Peptides
DM	:	Data Mining
DNA	:	Deoxyribonucleic Acid
DoH	:	Department of Health
DRAMP	:	Data Repository of Antimicrobial Peptides
E-value	:	Expectation value
FADE	:	Fast Atomic Density Evaluation
FBA	:	Fructose-1, 6-bisphosphate aldolase
fhbp	:	Factor H binding protein
Fn	:	Fibronectin
FN	:	False Negative
FP	:	False Positive
GAPDHs	:	Glyceraldehyde 3-phosphate dehydrogenases
GMI	:	Global Meningococcal Initiative
HCV	:	Hepatitis C virus
HD5	:	Human Defensin 5
HIV	:	Human Immunodeficiency Virus
HMM	:	Hidden Markov Model
HSPGs	:	Heparin Sulfate Proteoglycans
IL	:	Interleukin
IMD	:	Invasive Meningococcal Disease
IR	:	Information Retrieval

I-TASSER	:	Iterative Threading Assembly Refinement
KDD	:	Knowledge Discovery in Databases
KFC	:	Knowledge-based FADE and Contacts
LFD	:	Lateral Flow Device
LOL	:	Lipoprotein Outer membrane Localization
LPS	:	Lipopolysaccharide
MCC	:	Mathew's correlation coefficient
MD	:	Meningococcal disease
MFP	:	Membrane-Fusion Protein
MIC	:	Minimum Inhibitory Concentration
Msf	:	Meningococcal surface fibril
MspA	:	Meningococcal serine protease A
MW	:	Molecular Weight
NadA	:	Neisserial adhesin A
NCBI	:	National Center for Biotechnology Information
Neu5Ac	:	N-acetylneuraminic acid
Ngon	:	Neisseria gonorrhoea
NHBA	:	Neisserial heparin binding antigen
NhhA	:	Neisseria hia homolog A
NICD	:	National Institute for Communicable Diseases
NLM	:	National Library of Medicine
Nmen	:	Neisseria Meningitidis
NmenA	:	Neisseria meningitidis serogroup A
NmenC	:	Neisseria meningitidis serogroup C
NMR	:	Nuclear Magnetic Resonance

OM	:	Outer Membrane
OMPs	:	Outer Membrane Proteins
PCR	:	Polymerase Chain Reaction
PDB	:	Protein Data Bank
pI	:	Isoelectric point
POC	:	Point Of Care
POTRA	:	Polypeptide Transport Associated
PTMs	:	Post-Translational Modifications
RDT	:	Rapid Diagnostic Test
REGEX	:	Regular Expressions
RmpM	:	Reduction-modifiable protein M
RMSD	:	Root Mean Square Deviation
RNA	:	Ribonucleic Acid
RTX	:	Repeats-In-Toxins
SA	:	South Africa
SDM	:	Site-Directed Mutagenesis
SEC	:	Secretion
SLAM	:	Signalling Lymphocyte Activated Molecule
SLPs	:	Surface Lipoproteins
T1SS	:	Type one Secretion System
TAC	:	TaqMan Array Card
TAM	:	Translocation and Assembly Module
TAT	:	Twin Arginine Translocation
TbpA	:	Transferrin-binding proteins A
TbpB	:	Transferrin-binding proteins B

Tfp	:	Type IV pilus
TM-score	:	Template Modelling score
TN	:	True Negative
TP	:	True Positive
TPS	:	Two-Partner Secretion
TspA	:	T-cell stimulating protein A
USA	:	United States of America
Vn	:	Vitronectin
WHO	:	World Health Organization

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

LIST OF TABLES

LIST OF APPENDICES

Chapter 1

Literature Review

1.1. Introduction

Meningococcal disease (MD) was first described by Vieusseaux in 1805, when an outbreak swept through Geneva (Pace and Pollard, 2012). *Neisseria meningitides (N. meningitidis)* (the meningococcus), the causative agent of MD was identified in 1887 by the Austrian pathologist Anton Wiechselbaum (Rosenstein *et al*., 2001; Pace and Pollard, 2012). Meningococcal disease is an acute bacterial infection with the two common presentations of meningococcal infection, these being meningococcal meningitis (infection of the membranes that surround the brain and spinal cord) and meningococcemia (infection of the bloodstream). An infected individual may suffer one or both of these diseases. The most common presentations of invasive MD are meningitis and sepsis (Pace and Pollard, 2012). Over 500,000-1,200,000 invasive meningococcal diseases occur each year, with 50,000-135,000 deaths (Chang *et al.,* 2012; Jafri *et al*., 2013; Gabutti *et al.,* 2015). Young children, adolescents, and young adults suffer the greatest burden of disease from *N. meningitidis*. Children are particularly vulnerable to invasive MD because of their relative immune immaturity, in particular their relative under-responsiveness to pure polysaccharide antigens such as the meningococcal capsule (Nadel and Ninis, 2018).

Bacterial meningitis is a global public health concern, with several responsible etiologic agents that vary by age group and geographical area: *Escherichia coli*, *Haemophilus influenzae*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, group B *Streptococcus agalactiae*, *Staphylococcus aureus*, and *Listeria monocytogenes* (Oordt-Speets *et al.*, 2018).

In a meta-analysis study, it was determined that *S. pneumoniae* and *N. meningitidis* are the predominant pathogens in all age groups and all regions, as shown in Figure 1.1, accounting for 25.1–41.2% and 9.1–36.2% of bacterial meningitis cases, respectively (Oordt-Speets *et al.*, 2018) .

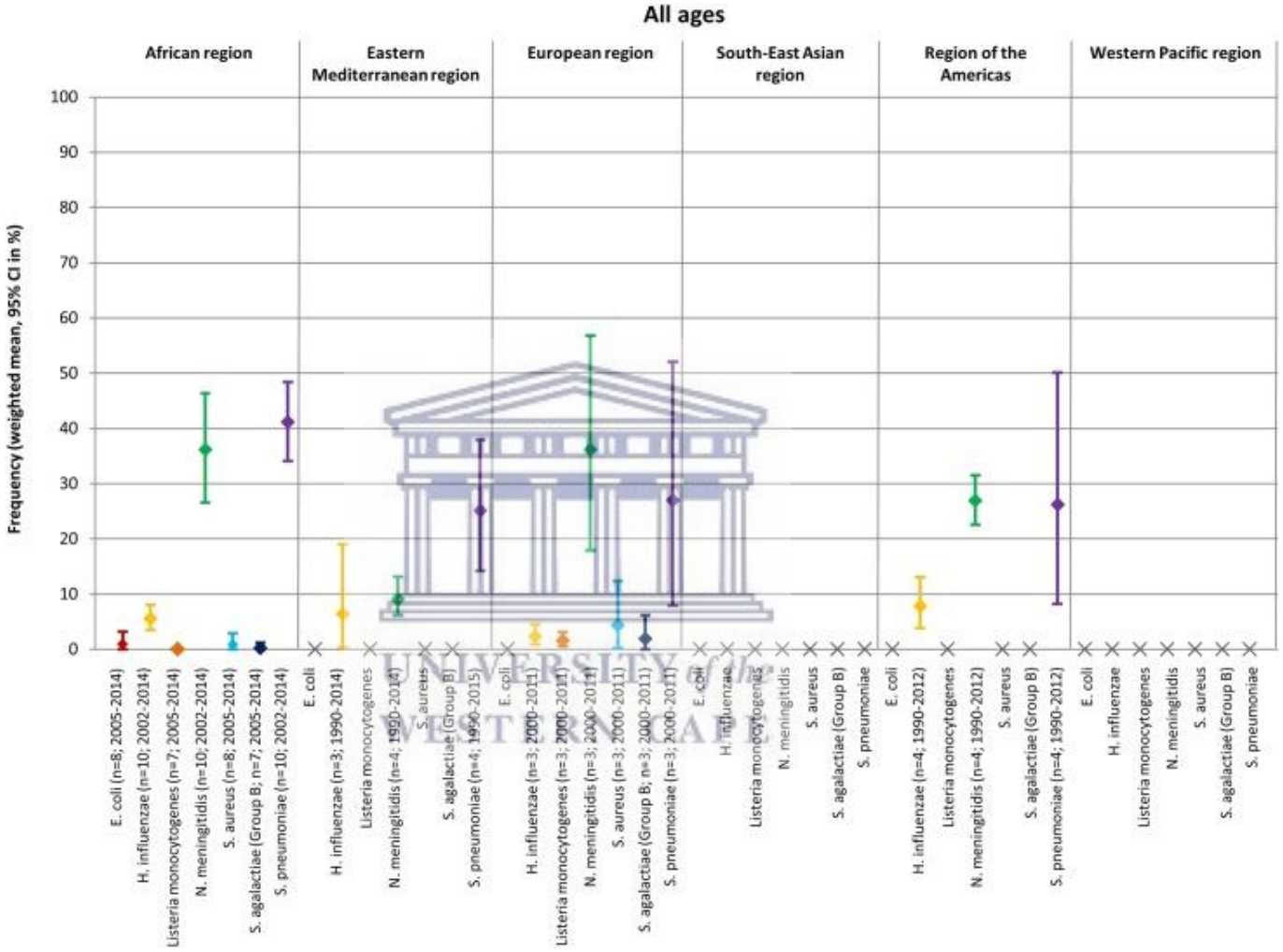


Figure 1.1: Frequency of seven pathogens that caused bacterial meningitis in all ages by geographic region (Extracted from Oordt-Speets *et al.*, 2018).

In the human respiratory tract, the only known reservoir of *N. meningitidis*, meningococci are exposed to human endogenous antimicrobial peptides (AMPs). Antimicrobial peptides are molecules that are part of the innate immune response and play an important role as a

host defense against microbial pathogens (Tzeng and Stephens, 2015). AMPs have a myriad of potential uses in the diagnosis and treatment of complex diseases such as cancer, where Bullard *et al.,* 2008 demonstrated Human Beta Defensin-1, an important component of the innate immune response, as a viable therapeutic agent for the treatment of late-stage prostate cancer. Terrin *et al.*, in 2011 investigated an innate immunity component, serum calprotectin as a biomarker of septicaemia in very low birth weight infants. Antimicrobial peptides can be exploited for the generation of novel antibiotics, as biomarkers in the diagnosis of inflammatory conditions, for the manipulation of the inflammatory process, wound healing, autoimmunity and in the combat of tumour cells (Mahlapuu *et al.,* 2016).

The advent of the use of automation equipment in molecular biology, coupled with the sequencing of the human genome has led to the discovery of a large number of AMPs associated with immune response. The sequences of these AMPs are curated in various databases and the use of high throughput screening via bioinformatics tools may help in understanding the function of peptides.

Computational tools are very useful for predicting novel antibacterial peptides, which could be used to design potent agents against bacterial pathogens. Extensive work has been done in the field of antibacterial peptides, describing their identification, characterization and mechanism of action. The information about these peptides has been collected and compiled; in major databases on AMPs (Lata *et al.*, 2007). Researchers are focused on *in silico* screening and modelling of novel AMPs as computational approaches can accelerate the process of antimicrobial discovery and design (Wang *et al.*, 2011; Li *et al.*, 2017).

1.2. Meningococcal Disease (MD)

Meningococcal disease is of major importance in public health due to its global distribution, epidemic potential, predominant disease burden in children and adolescents and fulminant clinical manifestations (Pace and Pollard, 2012; Lundbo *et al.,* 2015; Borrow *et al.,* 2017). The mild clinical presentation may progress to sudden onset of disease, multi-organ failure and death within hours (Nadel and Ninis, 2018). The clinical spectrum of invasive MD is diverse with meningitis and/or septicaemia being the commonest modes of presentation. Among survivors, disabling long-term sequelae can complicate meningococcal disease and result in potentially devastating effects on the quality of life of survivors, most of whom are infants, children and adolescents (Pace and Pollard, 2012). The presence of fever and cutaneous alterations petechia or purpura in an acutely ill patient should mandatorily evoke in the physician, the hypothesis of MD (Dwilow and Fanella, 2015; Batista *et al.,* 2017).

1.3. Pathogenesis of MD

The pathogenesis of meningococcal infection involves initial penetration of the nasopharyngeal mucosal epithelium, entry into the blood, and the development of bacteraemia, which occurs in the absence of host humoral immunity (Brandtzaeg and van Deuren, 2012). The meningococcus is carried in the human nasopharynx, asymptomatically, by five to ten percent of adults in non-epidemic periods (Tzeng and Stephens, 2000; Yazdankhah and Caugant, 2004; Yezli *et al.,* 2016). The interaction of *N. meningitidis* with human endothelial cells lining the blood vessels of the blood–cerebrospinal fluid barrier is a prerequisite for the development of meningitis (Coureuil *et al*., 2012; Doran *et al.,* 2016). *Neisseria meningitidis* may access the bloodstream, evolving by one of two primary mechanisms: (1) rapid bacterial multiplication, associated with development of marked

systemic inflammatory response, producing the typical clinical picture of meningococcemia; (2) slower reproduction of the agent, allowing for fixation and multiplication in the joints, the pericardium, and especially in the central nervous system, producing in the latter clinical picture of meningococcal meningitis (Batista *et al.*, 2017).

N. meningitidis possesses a variety of adaptive characteristics which enable it to avoid being killed by the immune system, such as the capsule, the lipopolysaccharide (LPS), groups of proteins that block the action of the AMPs, proteins that inhibit the complement system and components that prevent both the maturation and the perfect functioning of phagocytes (Gasparini *et al.*, 2012). The LPS (endotoxins) play a central role in the pathogenesis of the morbid condition, stimulating cells of the immune system (such as macrophages, monocytes and neutrophils) to release a series of inflammatory mediators: interleukin (IL) 1, IL-6, IL-8, IL-10, interferon-gamma and tumour necrosis factor alpha. These cytokines play a critical role in activation of multiple pathways. This increase in activity and events are directly responsible for the development of shock and multiple organ failure (Batista *et al.*, 2017).

1.4. Meningococcal colonization

The mechanisms that lead from colonisation to invasive disease, which occurs only in a small proportion of carriers, are still not completely understood but are thought to be as a result of meningococcal virulence factors, environmental conditions and host susceptibility (Li *et al.*, 2017; Nadel and Ninis, 2018). The blood-cerebrospinal fluid barrier physiologically protects the meningeal spaces from blood-borne bacterial pathogens. Few bacterial pathogens were able to reach the subarachnoid space and among those, *N. meningitidis* is the one that achieves this task the most constantly when present in the bloodstream. The small number of bacterial species capable of invading the meninges

suggests that specific virulence factors are required for bacteria to enter the subarachnoid space (Coureuil *et al*., 2012). Crossing the blood–brain barrier, the bacterium binds to endothelial cells of the cerebral microvasculature, and in the choroid plexus and capillaries of the encephalon at receptors. Once in the Cerebral Spinal Fluid (CSF), the pathogens begin replication and trigger inflammatory process in the subarachnoid space, with pathophysiological consequences (Batista *et al.*, 2017).

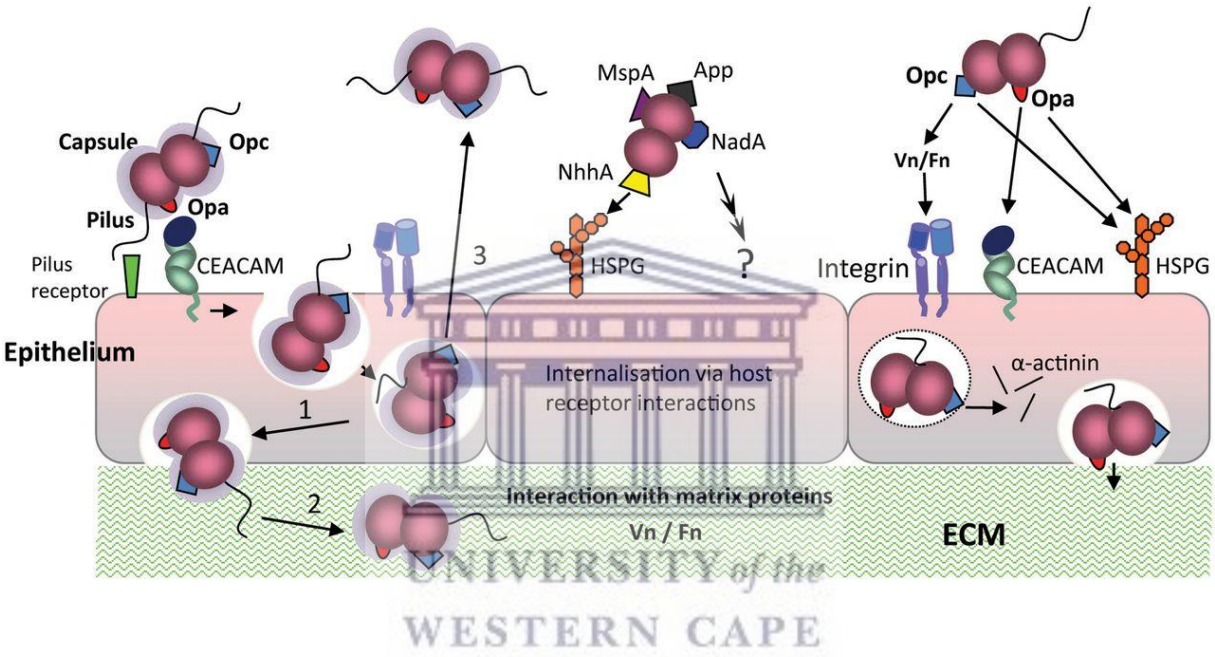


Figure 1.2: Schematic overview of meningococcal interactions at the epithelial barrier of the nasopharynx and the mode of barrier penetration (Adapted from Hill *et al*., 2010).

Although, the process of invasion and subsequent disease are not fully understood, a sequence of events must occur (Dwilow and Fanella, 2015). *N. meningitidis* enters the nasopharynx and attaches to non-ciliated epithelial cells, probably through the binding of the pili to the CD46 receptor (a membrane cofactor protein) and the subsequent binding of opacity-associated proteins, Opa and Opc, to the CD66e (carcinoembryonic antigen) and heparan sulfate proteoglycan receptors, respectively as shown in Figure 1.2 (Hill *et al*., 2010; Doran *et al.*, 2016). *Neisseria meningitidis* bind to members of the carcinoembryonic

antigen-related cell adhesion molecules (CEACAM) family of cell adhesion molecules and recognize specific glycoproteins in a lectin-like fashion. Binding of bacterial adhesins to specific host cell receptors may lead to a signal transduction resulting in tight bacterial attachment to or internalization by the host cells (Doran *et al.*, 2016).

The main means of adhesion of *N. meningitidis* to the host cells are Pili, type IV pili can retract through the bacterial cell wall, while the pilus tip remains attached to its target surface, allowing the so-called "twitching motility", a flagella-independent mode of motility important for efficient colonization of host surfaces (Ribet and Cossart, 2015). Opacity-associated proteins (Opa) and (Opc) are two proteins that make an important contribution to the process of adhesion to the cell. Porins A and B contribute to neisserial adhesion and penetration into the cells, and also inhibit the complement system. Factor H binding protein (fhbp) bind to factor H, allowing the bacteria to survive in the blood (Gasparini *et al.*, 2012). Once attached, these bacteria are much more resistant to shear stress and can start to proliferate leading to the formation of microcolonies (biofilms). Biofilm formation constitutes a protected mode of growth that allows bacteria to survive in a hostile environment (Ribet and Cossart, 2015).

Neisserial adhesin A (NadA) is a minor adhesin that is expressed by 50% of the pathogenic strains. NadA is known to be involved in cell adhesion and invasion and in the induction of proinflammatory cytokines. Neisserial heparin binding antigen (NHBA) binds heparin, thus increasing the resistance of the bacterium in the serum (Gasparini *et al.*, 2012).

Capsule down-modulation (or up-regulation of host receptors during inflammatory condition) allows interactions between outer-membrane proteins and their cognate host receptors. For example, Opa proteins may bind to the Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) and Heparin Sulfate Proteoglycans (HSPGs), and Opc proteins can

interact with HSPGs and, via vitronectin (Vn) and fibronectin (Fn), to their integrin receptors (Figure 1.2) (Hill *et al*., 2010).

Although some minor adhesins such as Neisseria hia homolog A (NhhA) have been shown to interact with HSPGs, the receptors targeted by meningococcal serine protease A (MspA), Adhesion and penetration protein (App) and NadA remain to be determined. Engagement of CEACAMs, integrins and HSPGs can result in meningococcal internalization by epithelial cells by triggering a variety of host cell signalling mechanisms. On crossing the epithelial barrier, meningococci are able to interact further with proteins of the extracellular matrix including Fn and Vn. Internalized bacteria may also migrate back to the apical surface for transmission to a new host (Hill *et al*., 2010).

1.5. Causative agent - *Neisseria Meningitidis* (*N. meningitidis*)

N. meningitidis are gram-negative, aerobic diplococci bacteria and a member of the bacterial family Neisseriaceae (Rouphael and Stephens, 2012). There are 13 serogroups of *N. meningitidis* based on different capsular polysaccharide structures, but only six serogroups (A, B, C, W-135, X, and Y) cause invasive meningococcal disease (Harrison *et al.,* 2013).

Meningococci are further classified on the basis of their class 1 outer membrane proteins (serosubtype), class 2 or 3 outer membrane proteins (serotype) and lipooligosaccharides (immunotype) (Rouphael and Stephens, 2012). Despite effective antibiotics and partially effective vaccines, *N. meningitidis* is the leading cause worldwide of meningitis and rapidly fatal sepsis usually in otherwise healthy individuals (Herwald and Egesten, 2011).

N. meningitidis penetrate host cellular barriers to initiate a local infection that can result in systemic spread associating in high-level bacteraemia and development of meningitis (Doran *et al.,* 2016). Humans are the only natural host of meningococcus. The bacterium is transmitted through respiratory droplets and close contact, with transmission increasing in crowded settings such as military camps, universities, and schools (Ali *et al.*, 2016).

N. meningitidis has outer and inner membranes surrounding a layer of peptidoglycan. The outer membrane contains important virulence factors including LPS and outer membrane proteins (OMPs), which function as porins (Figure 1.3). Other OMPs enhance adherence and invasion. Pili are also present on meningococci and play a key role in the process of adherence, colonization, and subsequent invasion. Recent *in vitro* work has identified a host cell surface receptor critical in the adhesion of pathogenic meningococci to endothelial cells (Dwilow and Fanella, 2015). *N. meningitidis* strains causing invasive disease and isolated from sterile sites such as the blood or the CSF are almost always encapsulated as seen in Figure 1.3. The capsule is essential for the survival of the organism in the blood as it provides resistance to antibody/complement-mediated killing and inhibits phagocytosis (Uria *et al*., 2008). The main meningococcal capsular polysaccharides associated with invasive disease are composed of sialic acid derivatives and expressing different capsular polysaccharides, suggesting that meningococcal clones can switch the type of capsule they express and can escape vaccine-induced or natural protective immunity by capsule switching (Rouphael and Stephens, 2012). The most successful meningococcal vaccines target its capsular polysaccharide, which is expressed on the surface of the bacterium. The bacterial pathogen *N. meningitidis* is able to escape the currently available capsule-based vaccines by undergoing capsule switching (Ji *et al.*, 2017). In *N. meningitidis,* N-acetylneuraminic acid (Neu5Ac) allows the meningococcus to become less visible to the host immune system. “Capsule switching” occurs due to genetic identity of parts of the capsule loci and is the result of horizontal exchange by transformation and recombination in the locus of serogroup specific capsule biosynthesis genes (Rouphael and Stephens, 2012).

Capsule switching may be an important virulence mechanism of meningococci and other encapsulated bacterial pathogens. Capsule switching is another mechanism of escape from vaccine-induced or natural protective immunity and a virulence mechanism shown by other encapsulated bacterial pathogens (e.g., *Streptococcus pneumoniae*) (Rouphael and Stephens, 2012).

The virulence factors of *N. meningitidis* include: the polysaccharide capsule: a structure that protects the etiologic agent from complement-mediated phagocytosis and lysis (Pizza and Rappuoli, 2015). The different sialic acid (serogroups B, C, Y, and W-135) and nonsialic acid (serogroup A) capsular polysaccharides expressed by *N. meningitidis* are major virulence factors. Lipopolysaccharide is an endotoxin responsible for toxic shock, meningococcal adhesion and activation of the innate immune system (Rouphael and Stephens, 2012). Adherence factor: type IV pilus, that binds to CD46 receptors, is a complex protein structure, located on the external plasma membrane, which plays an important role in pathogen adherence to epithelial and endothelial cells of *Homo sapiens* (Pizza and Rappuoli, 2015). External membrane proteins belonging to the porine class are believed to participate in adhesion and invasion of the host cell (Batista *et al.*, 2017).

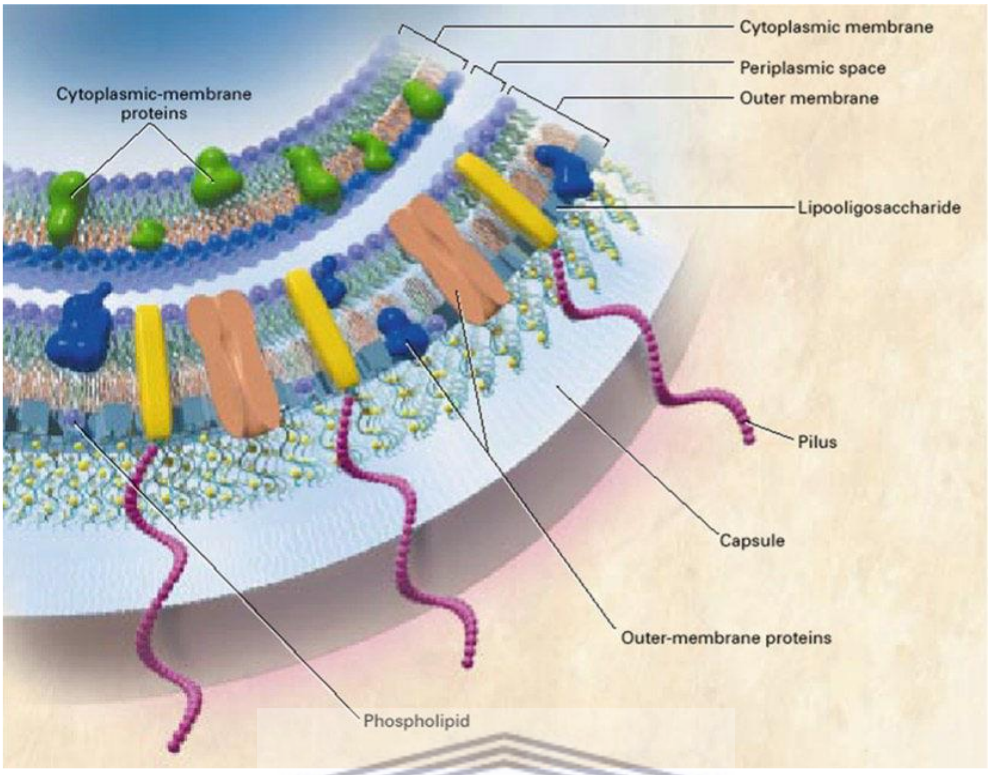


Figure 1.3: Cross-sectional view of the meningococcal cell membrane (Copyright© 2001 Massachusetts Medical Society. All rights reserved) (Rouphael and Stephens, 2012)

1.6. Epidemiology

The disease epidemiology caused by the different serogroups is constantly changing, both around the world and in different countries, due to selection pressure following introduction of effective vaccines and differences in antimicrobial usage (Nadel and Ninis, 2018). Different countries have different strains of the bacteria and of the 13 serogroups (strains) six exist having the potential to cause a major epidemic - A, B, C, X, Y and W135 responsible for virtually all cases of the disease in humans (Harrison *et al.*, 2013). Meningococcal disease can occur as endemic disease with sporadic cases, or epidemics with outbreaks of varying size and duration (Dwilow and Fanella, 2015). The disease is widespread around the globe and is known for its epidemical potential and high rates of lethality and morbidity (Batista *et al.*, 2017).

Meningitis epidemics in the African meningitis belt (Figure 1.4) constitute an enormous public health burden. Shortages of vaccine to control the new hyper-invasive strain portend a catastrophe with potential to affect as many as 34 million people in the region. By raising the alert now, World Health Organization (WHO) hopes to close the critical vaccine gap and keep this highly feared disease from sweeping across and potentially beyond West Africa (WHO, 2018).

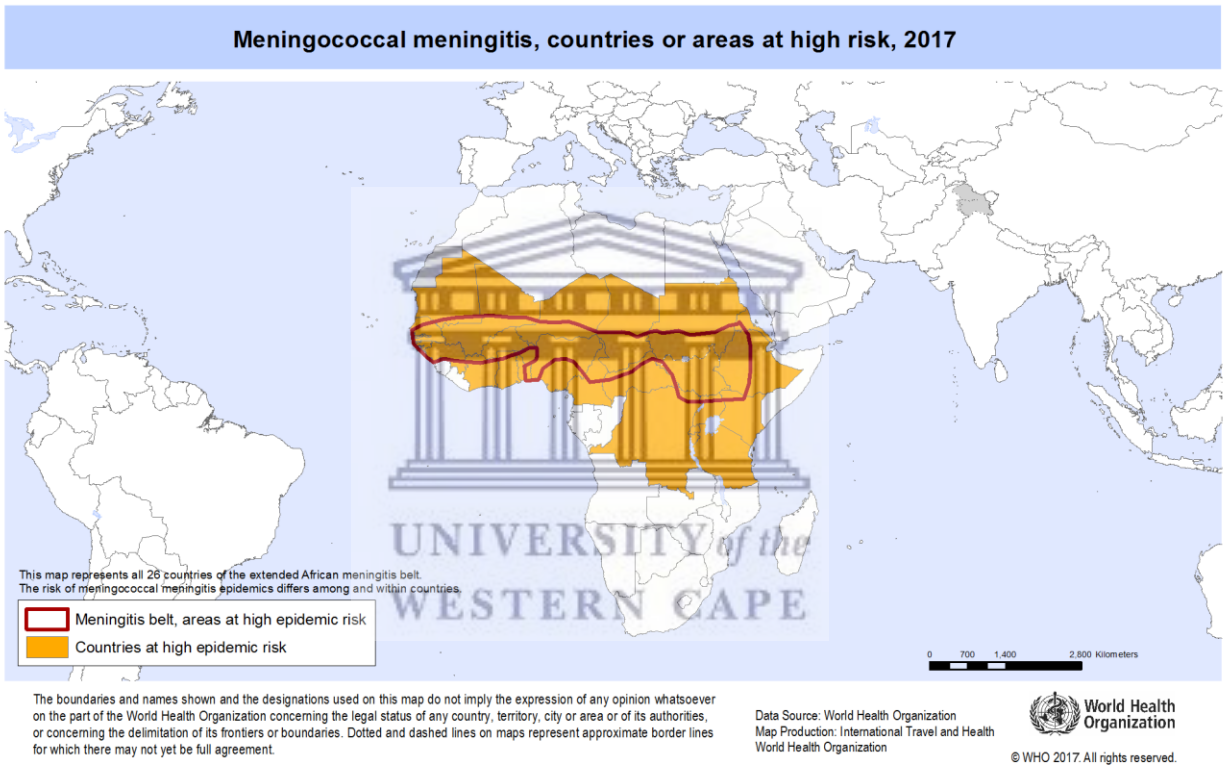


Figure 1.4: Geographical distribution and high epidemic risk of meningococcal meningitis

Since the introduction of the meningococcal A conjugate vaccine (*MenAfriVac*®), *N. meningitidis* serogroup A (*Nmen*A) cases have declined and *Nmen*A epidemics have been eliminated. However, in parallel, the proportion of cases and epidemics caused by other *N.*

meningitidis serogroups such as W, X, and C has increased (Chow *et al.*, 2016). More recently a *N. meningitidis* serogroup B protein-based vaccine has been developed and its use has been shown to limit outbreaks (Toneatto *et al.*, 2017).

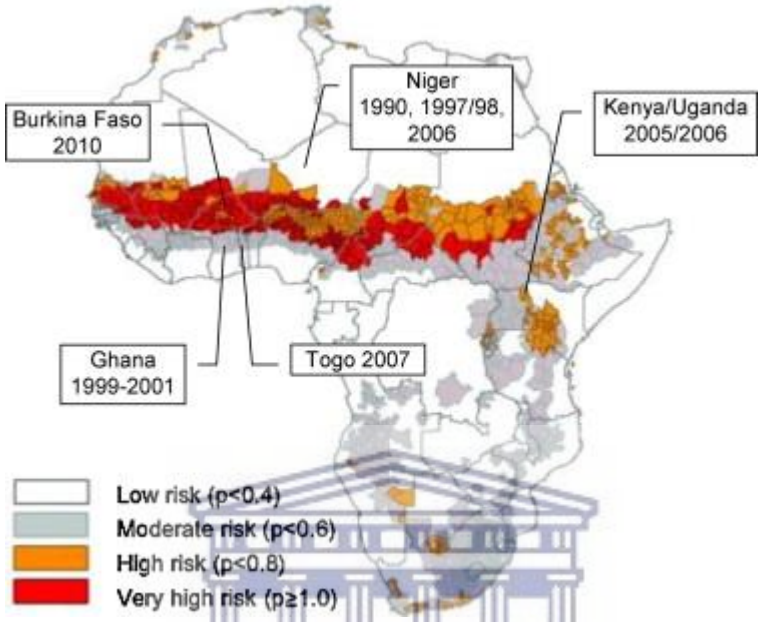


Figure 1.5: Countries in Africa with laboratory-confirmed outbreaks or larger clusters of cases of MD in the sub-Saharan meningitis belt involving serogroup X meningococci (Xie *et al.*, 2013).

In the meningitis belt of sub-Saharan Africa, seen in Figure 1.5, pandemics of meningococcal disease occur regularly and attack rates may exceed 800 cases per 100,000 populations per year. In some countries in this region, attack rates may be as high as 1 person in every 100 (Harrison *et al.*, 2009; Nadel and Ninis, 2018).

A new hyper-invasive strain of meningococcal meningitis serogroup C is circulating at the same time. Thus, an acute shortage of meningitis C-containing vaccine threatens to severely limit the ability to minimize the number of persons affected. The risk of imminent large-scale epidemics is dangerously high. In 2016 alone, the new strain of *N. meningitidis*

serogroup C (*Nmen*C) caused 18,000 cases in Nigeria and Niger. Attack rates can be very high (up to 670 cases/100,000), population immunity is low, the strain is already circulating in neighbouring countries of the African meningitis belt (Burkina Faso, Mali) and it showed a potential to spread outside the belt as observed in Liberia (WHO, 2017).

In the United States of America (USA), the outbreak rate is less than one case per 100,000 per year. During the 10-year period of 1998–2007 and according to the Active Bacterial Core surveillance (ABCs), the annual incidence decreased by 64.1%, from 0.92 cases per 100,000 in 1998 to 0.33 cases per 100,000 in 2007 with an average of 0.53 cases per 100,000 per year (Rouphael and Stephens, 2012). Infants aged less than one year have the highest incidence of MD (5.38 cases per 100,000) and the meningococcal vaccine is recommended in the USA for children 9 months to two years, who are at an increased risk of meningococcal disease. Since 2012, serogroups C, Y (since the mid-1990s) and B caused most disease in the USA and in Europe the attack rates (≥2 per 100,000 per year) have been higher than those observed in the USA. In the United Kingdom, rates of 5 per 100,000 per year prompted universal vaccination against serogroup C (Rouphael and Stephens, 2012).

In South Africa (SA) the pattern of MD is characterised by sporadic cases throughout the year with occasional small clusters and a definite seasonal increase in winter and early spring and outbreaks may especially occur in mines, correctional and detention facilities, academic institutions, and displaced communities. Vaccination is not given routinely but may be initiated in response to an outbreak (DoH, 2011; du Plessis *et al.,* 2012). The National Institute for Communicable Diseases (NICD) data on laboratory-confirmed cases indicate high incidences in the Gauteng and the Western Cape provinces. In the Western Cape serogroup B tends to be the most common serogroup and in areas of Gauteng and North West provinces, outbreaks occur with serogroup A, and to a lesser extent, serogroup C

predominating (Von Gottberg *et al.*, 2008; DoH, 2011; Jafri *et al.*, 2013). In South Africa, a quadrivalent protein-conjugated meningococcal vaccine (MCV4) is available, and provides protection against 75% of disease causing serogroups (Meiring *et al.*, 2017).

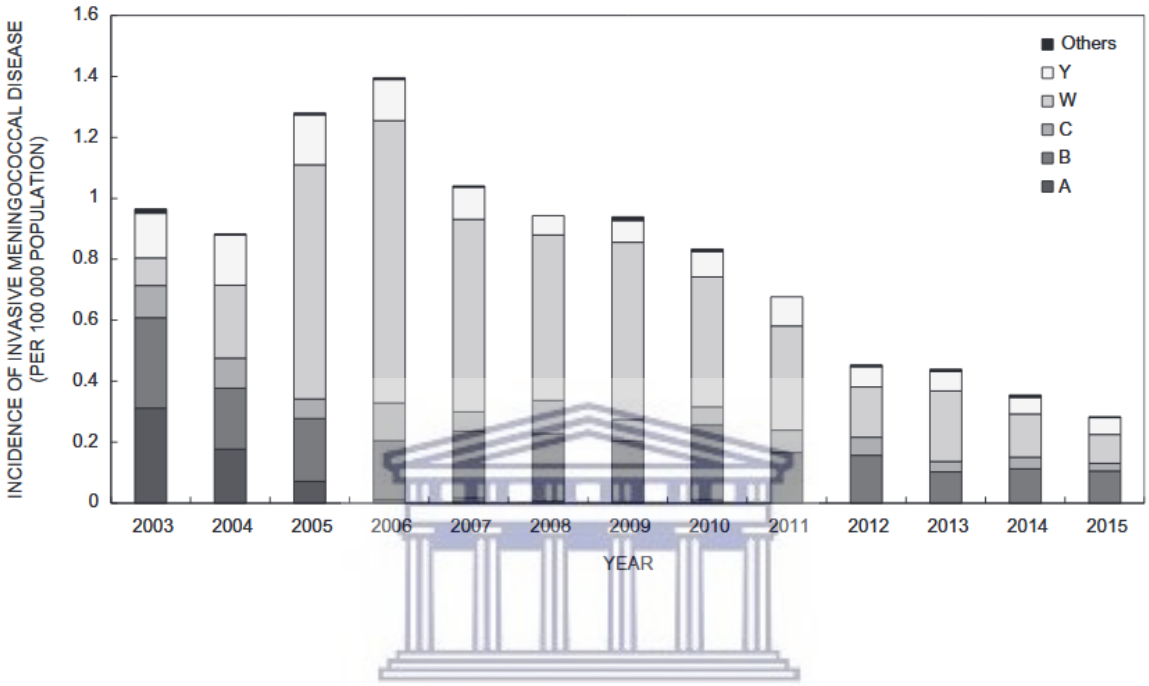


Figure 1.6: Incidence of invasive meningococcal disease by serogroup, South Africa, 2003-2015 (n = 5118) (Extracted from Meiring *et al.*, 2017).

The serogroups found to cause disease in South Africa (Figure 1.6), with the majority of disease caused by serogroup W, followed by serogroup B. Prior to 2005, serogroup A was South Africa's predominant disease causing serogroup but only a few cases have been detected over the last 3 years. Meningococcal case-loads are known to undergo alternate increases and decreases over periods of 5 to 10 years, therefore South Africa may possibly be on the verge of seeing an increase in meningococcal disease in the near future (Meiring *et al.*, 2017).

The burden of disease is highest in infants <1 year old, but the case fatality ratio is highest in adults and increases with age; in addition, human immunodeficiency virus (HIV) infection is a risk factor for contracting MD (Cohen *et al.*, 2010; Gianchecchi *et al.,* 2015). Several cases have also been reported in closed groups, such as Hajj pilgrims, adolescents, military and in students who live in dormitories (Gianchecchi *et al.,* 2015). The Global Meningococcal Initiative (GMI) was established in 2009 to promote the prevention of MD worldwide through education, research, international cooperation, and vaccination (Harrison *et al.*, 2011) and currently underline the need to increase the availability and quality of laboratory surveillance in order to understand the true burden of MD (Borrow *et al.,* 2017).

1.7. Antimicrobial peptides (AMPs)

Antimicrobial peptides (AMPs) are a diverse class of naturally occurring molecules that are produced as a first line of defense by all multicellular organisms ranging from prokaryotes to humans (Zhang and Gallo, 2016). Antimicrobial peptides are generally amphipathic, small in size (12–50 amino acids) and have at least two positive charges (as arginine and lysine residues). These chemical properties allow them to insert into the anionic cell wall and phospholipid membranes of microorganisms, disrupting them, resulting in bacterial killing (da Silva and Machado, 2012). The net positive charge of peptides is a common property (cationicity) of more than two thousand different anuran defense peptides discovered so far. Another common property of helix-forming AMPs is amphipathicity. One helix side is hydrophilic and charged, whilst the other helix side is hydrophobic and neutral (Juretić *et al.*, 2013). Antimicrobial peptides are gene-encoded, ribosomally synthesized polypeptides. They usually have common characteristics: small peptide size, strongly cationic (pI 8.9–10.7), heat-stable (100 ◦C, 15 min), no drug fastness and no effect on eukaryotic cells. Naturally occurring AMPs have been isolated and characterized from practically all-living organisms, ranging from prokaryotes to humans (Li *et al.,* 2012).

The discovery of AMPs dates back to 1939, when Dubos extracted an antimicrobial agent from a soil *Bacillus* strain. This extract was demonstrated to protect mice from pneumococci infection. In the following year, Hotchkiss and Dubos fractionated this extract and identified an AMP which was named gramicidin (Bahar and Ren, 2013). The first reported animal-originated AMP is defensin, which was isolated from rabbit leukocytes in 1956. In the following years, skin secretion of *Bombinin maxima* and lactoferrin from cow milk were both described (Bahar and Ren, 2013).

Lysozyme was the first reported human antimicrobial protein identified in 1922 from nasal mucus by Alexander Fleming. In the 1960s, the rise of multidrug-resistant microbial pathogens awakened the interest in AMPs as host defence molecules. Presently, more than 2,500 AMPs have been deposited in the Antimicrobial Peptide Database (APD) (Wang *et al.*, 2016). The ability of AMPs to kill bacteria usually depends upon their ability to interact with bacterial membranes or cell walls. Generally, AMPs exhibit a net positive charge and a high ratio of hydrophobic amino acids, allowing them to selectively bind to negatively charged bacterial membranes (Figure 1.7) (Zhang and Gallo, 2016). The growing problem of resistance to and overuse of conventional antibiotics has stimulated interest in the development of AMPs as the next generation anti-infective agents and as methods to more selectively combat pathogens (Zhang and Gallo, 2016).

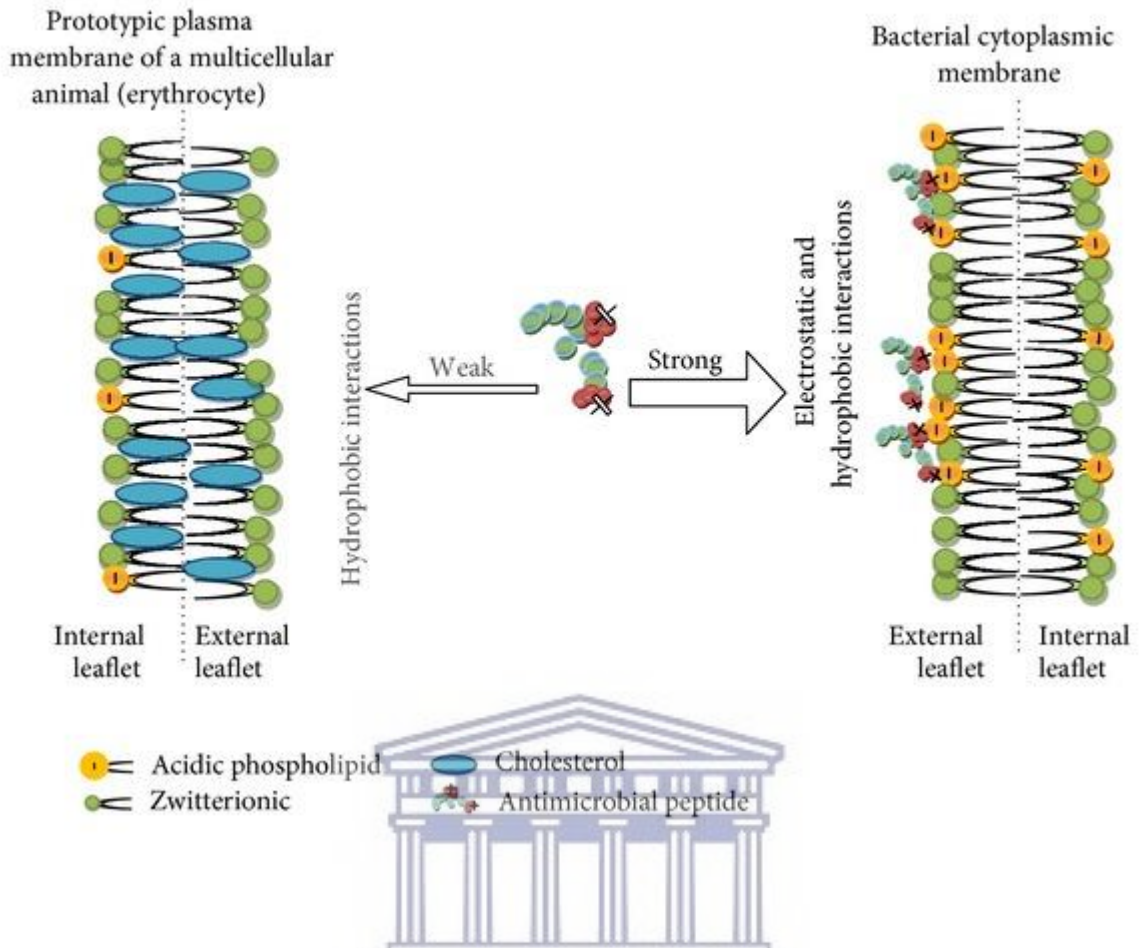


Figure 1.7: Membrane targeting of antimicrobial peptides and basis of their selectivity (Adapted from Ebenhan *et al.*, 2014).

Most AMPs have the prominent function to kill bacteria, which were used in recruitment, activation and/or maturation of inflammatory and immune cells and/or tissue repair. Most mammalian AMPs are produced by the inflammatory and epithelial cells as part of the host response to microbial invasion (Li *et al*., 2012). Many AMPs have multi-functions such as antibacterial, antifungal and anti-cancer activities. In addition to their antimicrobial action, AMPs have demonstrated diverse biologic effects, all of which participate in the control of infectious and inflammatory diseases, characteristics that make these peptides attractive as therapeutic tools. Further researches will advance knowledge within the field and highlight the potential of AMPs as therapeutic agents (Li *et al*., 2012).

Bacteria are commonly divided into two families, Gram-positive and Gram-negative, based on the differences in cell envelope structure. In Gram-positive bacteria, the cytoplasmic membrane is surrounded by a thick peptidoglycan layer, whereas the cytoplasmic membrane of Gram-negative bacteria is surrounded by a thin peptidoglycan layer as well as an outer membrane (Lin and Weibel, 2016). It is widely accepted that membrane interaction is a key factor for the direct antimicrobial activity of AMPs (Mahlapuu *et al.*, 2016).

Meningococci have evolved effective mechanisms to confer intrinsic and high levels of resistance to the action of AMPs (Tzeng and Stephens, 2015). AMP resistance mechanisms have been well-characterized in various Gram negative bacteria such as *N. meningitidis* to include (i) efflux pumps that export AMP from the periplasmic and intracellular compartments (ii) structural modifications of LPS and lipooligosaccharide to reduce interaction with AMPs; (iii) modulation of outer membrane permeability to limit entry and/or enhance excretion of AMPs and (iv) proteases that degrade AMPs (Tzeng and Stephens, 2015).

The cytoplasmic membranes of both Gram-positive and Gram-negative bacteria are rich in the phospholipids phosphatidylglycerol, cardiolipin, and phosphatidylserine, which have negatively charged head groups, highly attractive for positively charged AMPs (Ebenhan *et al*., 2014). LPSs in the outer membrane of Gram-negative bacteria provide additional electronegative charge to the bacterial surface (Lai and Gallo, 2009; Ebenhan *et al*., 2014). Interactions between AMPs and mammalian cell membrane occur mainly via hydrophobic interactions, which are relatively weak compared to the electrostatic interactions taking place between AMPs and bacterial membranes. Furthermore, mammalian cell membranes, unlike those of microbes, have a high content of cholesterol (Lai and Gallo, 2009). Importantly,

regardless of the exact mode of action and target site, the antibacterial activity of AMPs is dependent on the interaction with microbial membrane (Mahlapuu *et al*., 2016).

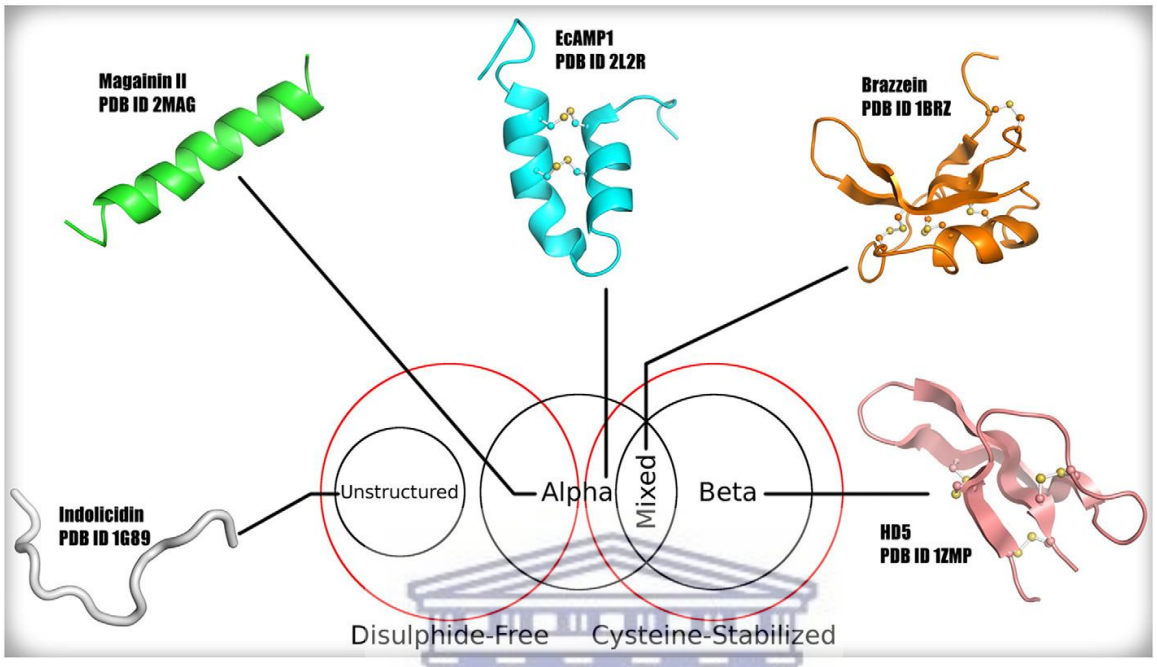


Figure 1.8: Classification and Structures of AMPs (Porto *et al.*, 2017).

There are two major classifications for AMPs, the first one being based on the structure and the second one on the presence or absence of disulphide bonds (Porto *et al.*, 2017). As seen in Figure 1.8, the structure and classification of AMPs: Indolicidin - unstructured peptides; magainin II - disulphide-free α-helical peptides; EcAMP1 - cysteine-stabilized α-helical peptides; brazzein - mixed α- and β-structure peptides and human defensin 5 - peptides composed only of β-strands. Disulphide bridges are represented as ball and stick (Porto *et al.*, 2017).

1.8. AMPs activity (mechanism)

The activities of antibacterial peptides are almost universally dependent upon interaction with the bacterial cell membrane. The first step in this interaction is the initial attraction between the peptide and the target cell, which is thought to occur through electrostatic bonding between the cationic peptide and negatively, charged components present on the outer bacterial envelope. The peptide can permeabilize the membrane and/or translocate across the membrane and into the cytoplasm without causing major membrane disruption (Jenssen *et al.*, 2006).

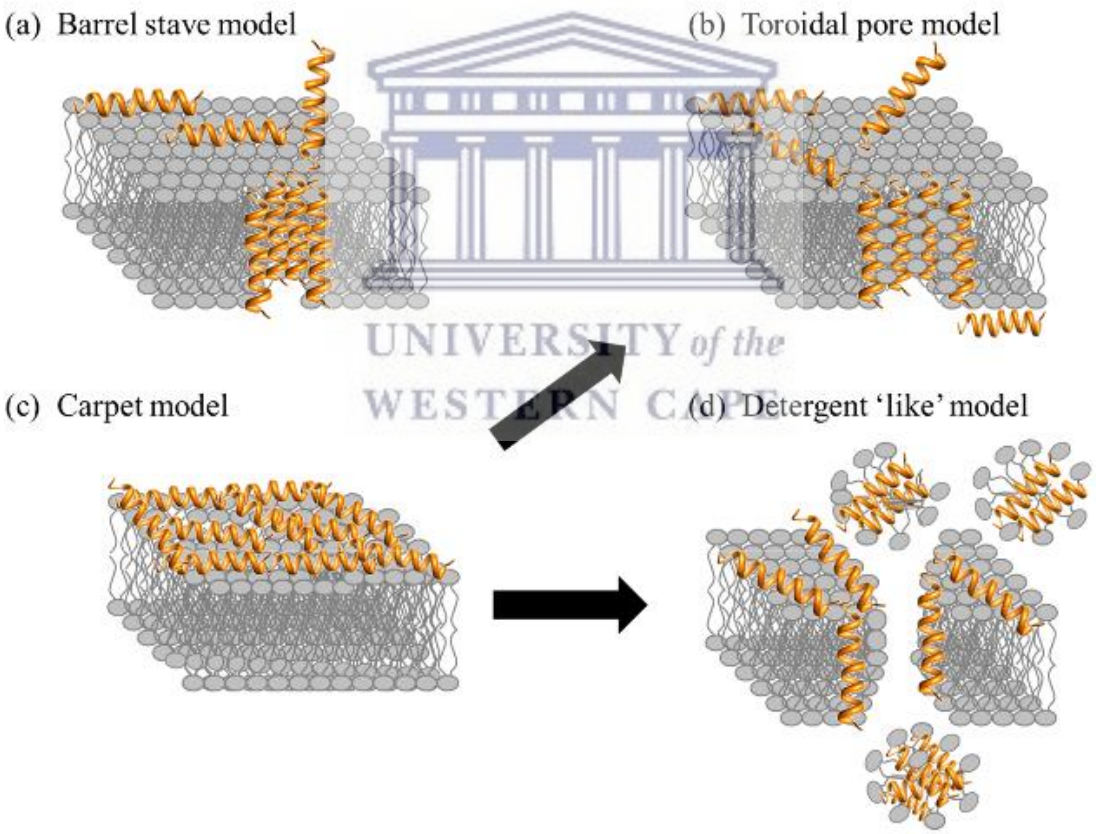


Figure 1.9: Mechanisms of action for AMPs in bacteria (Adapted from Kumar *et al.*, 2018).

The models of mechanisms of actions of AMPs can be classified under two broad categories: transmembrane pore and non-pore models. The transmembrane pore models can be further subdivided into the barrel-stave pore and toroidal pore models. In the barrel stave model, the AMPs are initially oriented parallel to the membrane but eventually insert perpendicularly in the lipid bilayer (Figure 1.9(a)). This promotes lateral peptide-peptide interactions where peptide amphipathic structure (minimum length of ~22 residues (α helical) and/or ~8 residues (β sheet)), is essential in this pore formation mechanism as the hydrophobic regions interact with the membrane lipids (Kumar *et al.*, 2018).

In the toroidal pore model, the peptides also insert perpendicularly in the lipid bilayer but specific peptide-peptide interactions are not present. The peptides induce a local curvature of the lipid bilayer with the pores partly formed by peptides and partly by the phospholipid head group (Figure 1.9(b)) (Kumar *et al.*, 2018).

In the "carpet model" (Figure 1.9(c)), accumulation of peptides on the membrane surface causes tension in the bilayer that ultimately leads to disruption of the membrane and formation of micelles (Jenssen *et al.*, 2006; Mahlapuu *et al.*, 2016; Kumar *et al.*, 2018). The final collapse of the membrane bilayer structure into micelles is also known as the detergent-like model (Figure 1.9(d)) (Kumar *et al.*, 2018). Importantly, regardless of the exact mode of action and target site, the antibacterial activity of AMPs is dependent on the interaction with microbial membrane (Mahlapuu *et al.*, 2016).

1.9. Bioinformatics – a search tool for AMPs discovery

When researchers realized that a gold mine of potentially new antibiotic classes exists in insects and amphibians (about 25 years ago), vigorous work started in isolating, characterizing and testing natural AMPs in parallel with exploration of chemical

modifications needed to make them more active against bacteria and less toxic for human cells (Juretić *et al.*, 2013). These methods of isolation and characterization included a peptide purification process and chromatography, which was both labour intensive and very costly.

Various bioinformatic methods have been developed for predicting as well as characterizing new AMPs. For example, APD has an embedded algorithm that can predict whether a new peptide has the potential to be classified as an antimicrobial based on certain known principles (Wang *et al.*, 2011). The computational approach to antimicrobial agent discovery and design encompasses genomics, molecular simulation and dynamics, molecular docking, structural and/or functional class prediction, and quantitative structure–activity relationships (Hammami and Fliss, 2010). Bioinformatic approaches are a valuable tool, since they have the advantages of fast sequence identification and low costs, if compared to the peptide purification process, allowing the rapid discovery of promising novel antimicrobial agents (Porto *et al*., 2017).

Despite their diverse origins, the majority of them have common biophysical parameters that are probably essential for activity, including small molecular size, cationicity and amphipathicity (Hammami and Fliss, 2010). Bioinformatics and wet-lab biology are interdependent and complement each other for the purposes of their own progress and for progress in antimicrobial discovery in the future (Hammami and Fliss, 2010). Algorithms for AMP identification uses tools such as profile Hidden Markov Models (HMMs).

1.9.1. Model Creation algorithm

Searching sequence databases is one of the most important applications in computational molecular biology (Eddy, 2011). HMMER is a software suite for protein sequence similarity searches using probabilistic methods. The HMMER website, provides access to the protein

homology search algorithms found in the HMMER software suite ((Finn *et al.*, 2015). The goal of the HMMER software is to make advanced probabilistic methods for sequence homology detection available in widely useful tools (Finn *et al.*, 2011).

1.9.2. What are Hidden Markov Models (HMMs)?

An HMM profile is a probabilistic model of a protein family multiple-sequence alignment, which uses position-specific scores to indicate the likelihood of each amino acid occurring in each position in the alignment (Eddy, 1998).

Profile HMMs are statistical models of multiple sequence alignments. They capture position-specific information about how conserved each column of the alignment is and which residues are likely. All the profile methods are more or less statistical descriptions of the consensus of a multiple sequence alignment. They use *position-specific* scores for amino acids (or nucleotides) and position specific penalties for opening and extending an insertion or deletion. Traditional pairwise alignment for example, FASTA uses position-*independent* scoring parameters. This property of profiles captures important information about the degree of conservation at various positions in the multiple alignment and the varying degree to which gaps and insertions are permitted (Eddy, 2003; Finn *et al.*, 2015).

1.9.3. Technical aspects

HMMs are a class of probabilistic models that are generally applicable to time series or linear sequences. An HMM describes a probability distribution over a potentially infinite number of sequences. Because a probability distribution must sum to one, the 'scores' that an HMM assigns to sequences are constrained. The probability of one sequence cannot be increased without decreasing the probability of one or more other sequences (Eddy, 1998).

Once an HMM is drawn, regardless of its complexity, the same standard dynamic programming algorithms can be used for aligning and scoring sequences with the model. The development of robust methods for automated sequence classification and annotation is imperative. Developing profile HMM methods hopefully can provide a second tier of solid, sensitive, statistically based analysis tools that complement current BLAST and FASTA analyses. The combination of powerful new HMM software and large sequence alignment databases of conserved protein domains should help make this hope a reality (Eddy, 1998).

Various steps are needed to create a profile HMM (i) Build a profile HMM model by aligning sequences (DNA or protein) from an input multiple alignment of many sequences to a common profile HMM. (ii) A HMM is then calibrated and empirically determines parameters that are used to make searches more sensitive, by calculating more accurate expectation value scores (E-values) and searched for a sequence database for matches to an HMM. (iii) Search a protein profile HMM against a protein sequence database and align sequences to an existing model (Eddy, 2003).

1.9.4. Examples of HMM use in discovery of a disease

The use of the HMMER algorithm is deemed an appropriate tool, which enables a more sophisticated search for novel peptides through proteome sequence scanning. Several AMPs discovered using HMM, have been proven to detect the HIV capsid protein p24 and a diagnostic kit has been developed employing these AMPs instead of the classically used antibodies (Williams *et al.*, 2016).

Recognition of highly divergent viral sequences is problematic and may be further complicated by the inherently high mutation rates of some viral types, especially RNA viruses. In these cases, increased sensitivity may be achieved by leveraging position-specific

information during the alignment process. A HMMER3-compatible profile HMMs was constructed from all the virally annotated proteins in RefSeq in an automated fashion using a custom-built bioinformatic pipeline and then tested the ability of these viral profile HMMs to accurately classify sequences as viral or non-viral (Skewes-Cox *et al.,* 2014).

Modelling Hepatitis C virus (HCV) to identify the virus mutation process is essential to its detection and predicting its evolution. El Nahas *et al.,* 2012, presented a model of HCV based on profile Hidden Markov Model architecture proposing to use it for detection of HCV in blood samples. Moreover, the HCV model will help in learning the mutation model of HCV and presents new therapeutic targets as well as genomic information for designing vaccine candidates. Tincho *et al.,* in 2016 demonstrated the ability of AMPs to inhibit HIV replication by the usage of profile Hidden Markov Model in design and discovery (Tincho *et al.,* 2016).

1.10. Diagnosis for MD

Gold standard culture diagnostic methods for meningococcal disease are too slow and frequently compromised by prior antibiotic treatment resulting in low sensitivity (Moore, 2018). However, there are challenges with the diagnosis, particularly in the developing world with underdeveloped microbiological services and without updated diagnostic methods such as polymerase chain reaction (PCR) (Nadel and Ninis, 2018).

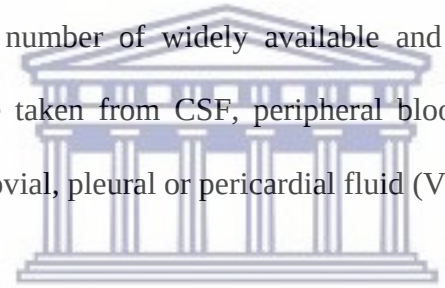
PCR is increasingly used for diagnosis of meningococcal meningitis including serogrouping and multilocus sequence typing and improved antigen and nucleic acid detection systems are enhancing the accuracy of epidemiological studies (Stephens *et al*., 2007). PCR-based methods require specific laboratory equipment and trained staff and cannot be used as a bedside method. PCR may not be sufficiently set to ensure countrywide surveillance,

especially in populations living in remote areas. Other tests, such as the currently available latex agglutination kits, require trained staff and an unbroken cold chain for storage and distribution of the kits (Agnememel *et al.,* 2015). PCR method may not be readily available for routine use and it may be more costly. Compared to conventional culture-based methods, PCR-based methods are not at risk of loss of viable organisms through early antibiotic treatment and sample processing (Vázquez *et al.,* 2016).

The development and application of sensitive quantitative Polymerase Chain Reaction (qPCR) assays has significantly improved laboratory detection rates and has reduced the time required to confirm invasive meningococcal disease. Currently, qPCR remains the preserve of a limited number of centralised reference laboratories who possess the necessary infrastructure, equipment and technical skills to routinely deliver an effective service. The time required to transport samples to centralised laboratories ultimately means that molecular detection of meningococci has little or no impact on patient management, whereby such testing merely confirms an initial clinical diagnosis and provides epidemiological data on circulating strains (Moore, 2018).

If diagnostic decisions are driven by clinical observations derived from hospital case-series, rather than the course of symptoms before admission, a diagnostic delay is inevitable. Rapid identification and treatment of disease complications such as shock, raised intracranial pressure and seizures, are vitally important to improve outcome (Nadel and Ninis, 2018). The initial non-specific manifestations mimic the symptoms of common viral infections and may create a diagnostic conundrum for the examining clinician. Early recognition and timely management of meningococcal septicaemia has reduced the mortality rate from 40% in the late 90s to 5–20% (Pace and Pollard, 2012).

Early diagnosis of MD is crucial to prevent further complications and improve the overall outcome for the affected individual thus early detection may further reduce individual morbidity and morbidity of MD. MD can lead to death in a healthy person within 6-12 hours of the first appearance of symptoms. The clinical diagnosis of meningococcal meningitis relies on the recognition of fever, rash, meningeal signs, and altered mental status, and is confirmed by pleocytosis, gram stain with or without culture of Cerebral Spinal Fluid (CSF), or blood or skin lesions. However, at early presentation, meningococcal disease and in particular, meningitis can have very non-specific signs and symptoms. The purpuric rash which aids diagnosis may be present early but often develops as the child is deteriorating (Stephens *et al*., 2007; Moore, 2018). The initial laboratory screening of a potential invasive MD case can incorporate a number of widely available and well-established laboratory evaluations; samples may be taken from CSF, peripheral blood or skin lesions or, when symptoms indicate, from synovial, pleural or pericardial fluid (Vázquez *et al.*, 2016).



N. meningitidis are aerobic agents, catalase-positive, oxidizing glucose and maltose with acid production and without gas formation. They grow well in chocolate and blood agar at temperatures between 35 ^{0}C and 37 ^{0}C, requiring an atmosphere of 5 %–10 % carbon dioxide (Rouphael and Stephens, 2012). Cerebral spinal fluid examination and culture lumbar puncture should be performed for suspected meningitis where no contraindications exist. The CSF should be kept as close to body temperature as possible whilst awaiting transport. The meningococcus is highly susceptible to heat, cold and direct sunlight. Tests requested on the CSF include: Protein and glucose determination (blood glucose should also be done), Direct microscopy (cell count and Gram stain) and culture and antibiotic susceptibility. There has been a renewal of interest in laboratory markers to help aid the diagnostic process and to help ensure that antibiotics are given at the earliest possible time (DoH, 2011).

A simple, rapid diagnostic assay that could assist with decision making in antibiotic prescription when MD is possibly present would be extremely helpful. Identification of *N. meningitidis* can be made on the basis of a cytological examination of the CSF, specific colony morphology on blood and/or chocolate agar, staining properties on a Gram stain, or by detection of specific antigens in the CSF by a latex agglutination test or using a rapid diagnostic test (RDT) (Organization, 2011). Spinal fluid cultures are positive in up to 90 % of cases later diagnosed as meningitis and blood cultures have only been reported to be positive in 40–75 % of cases suspicious for meningococcemia (Pollard and Finn, 2012).

The Pasteur Institute in Paris and the Centre deRecherche Me dicale et Sanitaire (CERMES) in Niamey, Niger developed a new dipstick RDT for the diagnosis of *N. meningitidis* serogroups A, W135, C and Y without prior sample preparation. This test exists as a duplex of two dipsticks (RDT1 and RDT2) in which RDT1 detects *N. meningitidis* serogroups A and W135/Y, while RDT2 detects *N. meningitidis* serogroups C and Y. An algorithm based on the results of the two dipsticks thus allows the detection of *N. meningitidis* serogroups A, C, W135 or Y (Rose *et al.*, 2009). The Pastorex test also carries certain constraints such as, restrictions of test reagents temperatures which must be kept between +4°C and +8°C. Manufacturer recommends on sample handling such as heating and centrifuging prior to using the test must be adhered to (Rose *et al.*, 2009).

In a study by Uadiale *et al.*, 2016, they used the Pastorex to demonstrate the good performance in detecting *N. meningitidis* serogroup C under field conditions, although they detected a limitation whereby the prepositioning Pastorex at peripheral health facilities during non-epidemic periods was constrained by a short shelf-life of 1 month after the kit is opened. Thus there is need for development of RDTs that are cheaper and with less challenging requirements for storage and usage (Uadiale *et al.*, 2016).

The dipstick RDT is better at correctly ruling out disease (as it has a lower false negative rate), while the Pastorex H test is better at confirming disease (having a lower false positive rate). However, the performance of the dipstick RDT on diluted whole blood samples was not satisfactory, as specificity was very low (57%). The sensitivity of this test on diluted whole blood, however, was similar to that of the Pastorex H test conducted on unprepared CSF (73% vs 69%) (Rose *et al.*, 2009). Latex agglutination testing to identify meningococcal antigen from sterile fluids were more widely used but its limitations included poor overall sensitivity and specificity, poor sensitivity for certain serogroups and had the inability to identify certain serogroups thus preventing it as being a routine test (Dwilow and Fanella, 2015).

Test kits are designed to be used by non-specialized staff with minimal training in basic facilities. Thus far, tests have been developed for capsular groups A, C, W, Y (Chanteau *et al.*, 2006) and X (Agnememel *et al.*, 2015).

Recently, the cryptococcal antigen lateral flow assay (CrAg LFA) was included as a resource for diagnosis. Unlike other tests, the CrAg LFA is a dipstick immunochromatographic assay, in a format similar to the home pregnancy test, and requires little or no lab infrastructure. This test meets all of the WHO ASSURED (Affordable, Sensitive, Specific, User friendly, Rapid/robust, Equipment-free, and Delivered) criteria. CrAg LFA in serum, plasma, whole blood, or cerebrospinal fluid is useful for the diagnosis of disease caused by *Cryptococcus* species (Vidal and Boulware, 2015).

Lateral flow assays will likely become more commonplace in all of infectious disease diagnostics, including meningitis. An unresolved question is the degree to which new diagnostics will become available in low-income countries (Bahr and Boulware, 2014).

In a recent study by Williams *et al.,* 2015, they reported 100% agreement between whole blood, serum and plasma LFA results demonstrated that fingerstick CrAg is a reliable bedside diagnostic test.

1.11. Biomarkers currently used for diagnosis

The inflammatory response elicited by *N. meningitidis* is considered to be the major factor responsible for the outcome of MD (Beran *et al.,* 2009). Cytokines and chemokines that were concentrated in the CSF were also detectable in the serum in the majority of MD patients. It is well known that IL-6 and IL-8 can be used in the laboratory diagnostics of bacterial meningitis and sepsis (Kleine *et al.,* 2003).

The findings of Beran *et al*., 2009 also indicated the potential use of IL-1ra, MCP-1, and MIP-1β for the laboratory diagnostics of sepsis, as well as bacterial meningitis tests are more laborious, expensive, time consuming, and less sensitive to distinguish bacterial meningitis treated with antibiotics from bacterial meningitis. As specific tests, e.g. bacterial culture and Gram stains (both are time consuming in performance), and latex agglutination, exhibited 70–90% sensitivity; no statistical CSF test proved to be fully reliable in distinguishing bacterial from viral meningitis thus far. Therefore, the results available with old classical markers of the current routine, e.g. lactate, leukocyte and granulocyte counts, Blood Brain Barrier (BBB) markers, glucose, must be reviewed in the context of clinical findings to diagnose bacterial meningitis reliably. Combinations of the CSF lactate test with CSF IL-6 test or with CSF leukocyte counts >800 M/l improved the lactate test. Moreover, the test with both the classical markers also proved to distinguish BM from intracerebral bleedings, stroke, Central Nervous System (CNS) tumours, and seizures as well (Beran *et al.,* 2009).

C-reactive protein (CRP) is one of the acute phase reactants produced by hepatocytes in response to inflammation. Its production is stimulated by IL-1 and IL-6 released from macrophages after activation by tissue damage, infection (bacterial, viral, fungal) or inflammation. An increase in CRP concentration in response to generalized inflammation occurs gradually during the first 12 h, reaching a maximum level at 48 to 72 h. Serum concentration of CRP higher than 80 mg/dl is a useful marker of bacterial etiology of meningitis (Gowin *et al.,* 2016). The tests currently available are either non-specific, e.g. determination of leukocyte count and CRP, or the results are not available soon enough to influence early treatment, e.g. PCR and blood culture. The level of procalcitonin in serum is increasingly being considered as a possible diagnostic tool for bacterial infections. The serum level of procalcitonin, which increases rapidly after endotoxin challenge, has been considered to be a relatively bacterio-specific marker with a high sensitivity for systemic bacterial sepsis and meningitis (Mills *et al.,* 2006).

1.12. Conclusion

One of the problems facing clinicians is that the spectrum of MD can present with non-specific signs, particularly in the early stages or in young children. Finding mechanisms to distinguish those with minor illnesses from those with more serious and potentially life threatening disease such as meningitis is crucial. Patients with these diseases can also deteriorate quickly. A major problem for MD is the diagnosis difficulties and experts conclude that with an early intervention the patient can survive. Recently there has been a renewal of interest in laboratory markers to help aid the diagnostic process and to help ensure that antibiotics are given at the earliest possible time.

Current targeting conventional markers (CRP, white blood cell, tumour necrosis factor-α, ILs, etc.) are non-specific for diagnosing sepsis. Procalcitonin, a member of the calcitonin

super family could be a critical tool for the diagnosis of sepsis. But to distinguish between bacterial versus viral infections, procalcitonin alone may not be effective (Vijayan *et al.*, 2017). Diagnosis relies on identifying typical features, which are not present in every case, particularly in the early stages. Procalcitonin appears to be better than white cell count and CRP, but still lacks sufficient accuracy to predict this infection (Bourke *et al.*, 2010).

Antimicrobial peptides play an important role as a host defense against microbial pathogens and are key components of the human innate immune response. It is therefore not surprising that meningococci have evolved effective mechanisms to confer intrinsic and high levels of resistance to the action of antibiotics and thus the importance of AMPs as new agents for detection and even so as therapeutics (Tzeng and Stephens, 2015).

Combining experimental data with computational biology will ultimately enable better understanding of antimicrobial agent–target interaction and the ability to manipulate biological systems more efficiently (Hammami and Fliss, 2010).

Over the past few decades, natural AMPs have been an active area of research and have shown a lowered likelihood for bacteria to form resistance compared to many conventional drugs (Veltri *et al.*, 2018).

Hoyos-Nogués *et al.*, 2018 has shown that AMPs are also capable of achieving good levels of specificity, discriminating between pathogenic and non-pathogenic bacteria or between Gram-positive and Gram-negative species, by using biosensors for the early detection of bacterial infections.

In another study by Mannoor *et al.*, 2010, they reported a simulated “water-sampling” chip, consisting of a microfluidic flow cell integrated onto the hybrid sensor, which demonstrates real-time on-chip monitoring of the interaction of *E. coli* cells with the antimicrobial

peptides, thereby opening exciting avenues in both fundamental studies of the interactions of bacteria with antimicrobial peptides, as well as the practical use of these devices as portable pathogen detectors. Real-time sensing results demonstrate the capability of the relatively simple impedance-based transduction architecture to directly detect bacteria.

A new much speedier method to rapidly identify *N. meningitidis* is required due to the urgent medical attention and deteriotion of the patient with MD. Early diagnosis and treatment may also help to halt the spread of the infection during epidemics and this can be done if diagnostic test are point-of-care and available at hospital beds.

Therefore an introduction of AMPs to attach to *N. meningitidis*, would provide a solution to this current diagnostic problem. This *in silico* research should potentially identify AMPs which are able to bind to the bacteria and detect or inhibit proliferation. Thus based on these predictions it would likely improve alternative approaches to diagnosis of MD.

The AMPs will provide a new way of diagnosing patients with MD. The AMPs will not be used as a therapeutic method since the bacterium has resistance mechanisms against the action of AMPs. Since the AMPs has to physically interact with the bacteria to exert its activity, this interaction can be exploited for diagnostic purposes with the use of a reporter molecule such as gold nano-particles as has been demonstrated by Williams *et al.*, 2016, where AMPs where employed for the detection of HIV.

1.13. Aims

The aim of this research work is to identify novel AMPs with activity against *N. meningitidis* for diagnosis of MD using *in silico* model creation.

The overall aim will be reached through these objectives:

Objectives

Objective 1: Collect all AMPs with antibacterial properties within the various databases and verify that they are experimentally validated as having activity against *N. meningitidis*. Cross reference between databases to create a final list of anti-*N. meningitidis* AMPs.

Objective 2: Extract unique identifiers and amino acid sequences of the aforementioned AMPs to create a predictive model as to identify novel AMPs against the bacteria *N. meningitidis.*

Objective 3: Construct Hidden Markov Models (HMM) models to identify the novel antimicrobial peptides from various genome sequences.

Objective 4: Identify *N. meningitidis* receptors through literature mining to which the newly identified anti-*N. meningitidis* AMPs will interact with.

Objective 5: Structural modelling of these AMPs as well as the *N. meningitidis* receptors using I-TASSER.

Objective 6: Site-directed mutagenesis (SDM) of amino acids within these AMPs to increase their binding affinity to the *N. meningitidis* receptors.

Objective 7: Interaction studies of these AMPs to the *N. meningitidis* receptors using PatchDock.

Chapter 2

Generation of a putative AMP list for detection of *N. meningitidis* using a bioinformatics approach

2.1. Introduction

MD is of major importance in public health due to its global distribution, epidemic potential, predominant disease burden in children and adolescents, fulminant clinical manifestations and high rates of lethality and morbidity (Pace and Pollard, 2012; Batista *et al.,* 2017). MD is an acute infection with the two common presentations of meningococcal infection being meningococcal meningitis and meningococcemia (sepsis) (Pace and Pollard, 2012).

Of the diagnostic markers introduced into clinical practice, procalcitonin appears to have the best attributes, although not specific for MD. The disease is hard to diagnose because initial symptoms are similar to those of influenza. Isolating and identifying this pathogen using conventional biochemical methods require 48-72 h. While simple PCR-based tests that are specific for the genus Neisseria and the species *N. meningitidis* have been developed (Lansac *et al*., 2000). Rapid, accurate diagnosis is essential for optimal management of patients. It is becoming increasingly difficult to confirm the diagnosis of meningococcal infection by conventional microscopy and culturing techniques (Bryant *et al*., 2004). Other limitations of conventional diagnostic methods include the delay before cultures become positive and the poor sensitivity and specificity of rapid antigen and antibody tests (Bryant *et al*., 2004).

AMPs are molecules small in size (12–50 amino acids) that provide protection against environmental pathogens, acting against a large number of microorganisms, including bacteria, fungi, yeast and virus. Their production is a major component of innate immunity against infection. In the human respiratory tract, the only known reservoir of *N. meningitidis*, meningococci are exposed to human endogenous AMPs (Tzeng and Stephens, 2015). AMPs

are gene-encoded, ribosomally synthesized polypeptides with common characteristics: strongly cationic (Isoelectric point (pI) 8.9–10.7), heat-stable (100 ◦C, 15 min), no drug fastness and no effect on eukaryotic cells. Naturally occurring AMPs have been isolated and characterized from practically all-living organisms, ranging from prokaryotes to humans (Li *et al.,* 2012).

The growing problem of resistance to and overuse of conventional antibiotics has stimulated interest in the development of AMPs as the next generation anti infectives and as methods to more selectively combat pathogens (Zhang and Gallo, 2016). AMPs have a myriad of potential uses in the diagnosis and treatment of complex diseases. AMPs can be exploited for the generation of novel antibiotics, as biomarkers in the diagnosis of inflammatory conditions, for the manipulation of the inflammatory process, wound healing, autoimmunity and in the combat of tumour cells (da Silva and Machado, 2012).

In this study, an *in silico* approach was sought to identify potential AMPs as identifiers of *N. meningitidis* receptors in order to develop a diagnostic device with high sensitivity, efficacy and accuracy and with lower sample requirement and minimal result interpretation for *N. meningitidis* diagnostics.

The AMPs identified will not be used as a therapeutic method since the bacterium has resistance mechanisms against the action of AMPs. Since the AMPs has to physically interact with the bacteria to exert its activity, this interaction can be exploited for diagnostic purposes with the use of a reporter molecule such as gold nano-particles as has been demonstrated by Williams *et al.,* 2016, where AMPs were employed for the identification of HIV.

2.2. Computational approaches used for the discovery of putative AMPs

The organization of data into a unified resource (databases) ensures data quality and integrity and brings considerable benefits in terms of synergy and efficiency; however, there are

several antimicrobial databases of very similar or identical scope that act in competition rather than in a network. Cross-linking between these would make their use more efficient, and researchers would benefit from such synergy (Hammami and Fliss, 2010). The large variety of AMP databases developed to date is characterized by a substantial overlap of data and similarity of sequences with the majority of databases having their own set of unique sequences (Aguilera-Mendoza *et al.*, 2015). Hence the reason and conclusion to include more than one database in this research project for searching and extracting AMPs.

The identification of AMPs from databases has gained attention as a branch of structural genomics and bioinformatics and several approaches have been applied for the identification of AMPs from databases, including local alignments, activity prediction by machine learning methods and also three-dimensional (3D) structure predictions (see Figure 2.1.). For biotechnology, these approaches are a valuable tool, since they have the advantages of fast sequence identification and low costs, if compared to the peptide purification process, allowing the rapid discovery of promising novel antimicrobial agents (Porto *et al.*, 2017).

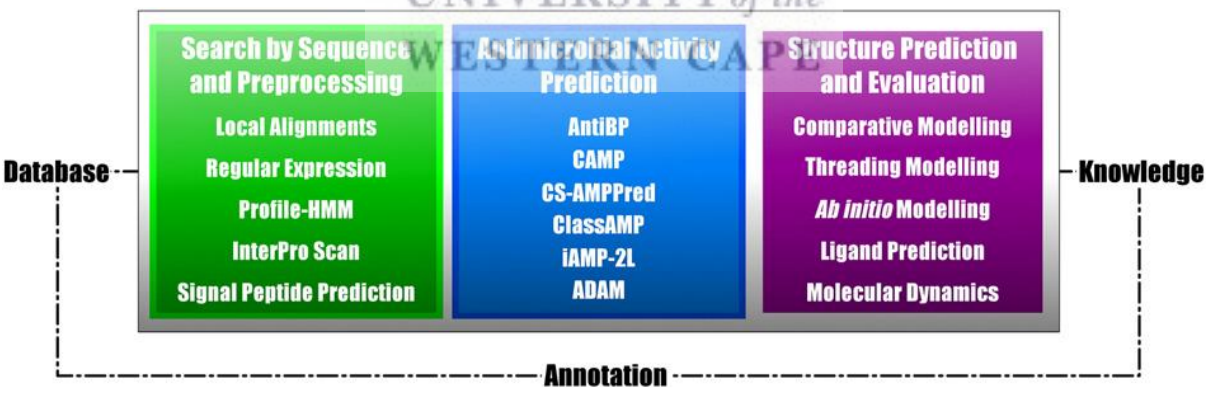


Figure 2.1: Current methods for mining databases for generating knowledge and useful data for sequence annotation (Porto *et al.*, 2017).

The classic method for discovering natural AMPs is to isolate them from a natural source chromatographically and in terms of biological source, amino acid sequence, 3D structure

and antimicrobial activity, this method shaped the current views on AMPs. Identification via chromatography is demanding and costly and thus a need to discover AMPs by different approaches: computer-based prediction and design (Wang, 2017).

2.2.1. Data mining

Data mining (DM) refers to extracting or "mining" knowledge from large amounts of data by finding new interesting patterns and relationships within this data. It is defined as "the process of discovering meaningful new correlations, patterns and trends by digging into large amounts of data stored in warehouses" and is also sometimes referred to as Knowledge Discovery in Databases (KDD). DM approaches seem ideally suited for bioinformatics, since it is data-rich (Raza, 2012) and the bioinformatics principle approach is to compare and group the data according to biologically meaningful similarities and then, based on this, analysing one type of data to infer and understand the observations for another type of data (Kamble and Khairkar, 2017). DM employs a wide spectrum of well-established statistical and machine learning techniques such as neural networks and other advanced algorithms and computational techniques to derive hidden meaningful correlations, patterns and trends from various biological databases (Ioannou *et al.*, 2014).

2.2.2. Biological databases

Biological databases (BD) are libraries of life sciences information, collected from scientific experiments, published literature, high-throughput experiment technology, and computational analysis (Attwood *et al.*, 2011). BD offer scientists the opportunity to access a wide variety of biologically relevant data, including the genomic sequences of an increasingly broad range of organisms (Baxevanis and Bateman, 2015). The aim of BD is to facilitate data retrieval and visualization as well as provide web application programming interfaces (APIs) for computers to exchange and integrate data from various database resources in an automated

manner, which is a fundamentally essential task in bioinformatics. BD integrate enormous amounts of omics data, serving as crucially important resources and becoming increasingly indispensable for scientists from wet-lab biologists to *in silico* bioinformaticians (Zou *et al.*, 2015).

Several databases hosting AMPs were employed within this study, which is summarized in Table 2.1 and fully discussed in the subsequent paragraphs.

2.2.2.1. Antimicrobial Peptide Database (APD)

In the past, extensive work has been done in the field of antibacterial peptides, describing their identification, characterization and mechanism of action. The information about these peptides has been collected and compiled in databases on AMPs with APD being one such database and consists of detailed information for AMPs (Lata *et al.*, 2007).

The Antimicrobial Peptide Database (APD) (http://aps.unmc.edu/AP/) web-accessible program provides interactive interfaces for peptide query, prediction and design. It also provides statistical data for a select group of, or all the peptides in the database. Peptide information can be searched using keywords such as peptide name, ID, length, net charge, hydrophobic percentage, key residue, unique sequence motif, structure and activity (Wang and Wang, 2004). The initial APD was established in 2003 and a brief description of the database was published in an issue of Nucleic Acid Research in 2004 in order to promote the research, education and information retrieval and knowledge discovery in the antimicrobial peptide field. The new version which has been regularly updated and further expanded from APD2 (2009 version) into APD3, currently focuses on natural antimicrobial peptides (AMPs) with defined sequence and activity (Wang *et al.*, 2015).

Table 2.1: The databases used for extraction of AMPs with activity against *N. meningitidis*

Database	Size (Sept 2018)	Site URL	Description	Reference
APD	3013 AMPs	http://aps.unmc.edu/AP/	Stores AMPs from all biological sources. Manually collected and curated from the literature (via PubMed, PDB, Google and Swiss-Prot), including some synthetic peptides.	Wang and Wang 2004; Wang *et al.,* 2008; Wang *et al.,* 2016
CAMP	8164 AMPs	http://www.bicnirrh.res.in/antimicrobial	Sequences and structural information of AMPs were retrieved from protein databases such as NCBI, UniProt and PDB. CAMP is sectioned into sequence, structure and patent databases. Experimentally validated peptides and predicted ones.	Thomas *et al.,* 2009; Waghu *et al.,* 2014; Waghu *et al.,* 2015
DRAMP	17629 AMPs	http://dramp.cpu-bioinfor.org	This database is divided into three datasets: general, patent, and clinical AMPs. It harbours diverse annotations, including sequence, structure, physicochemical, patent, and clinical information.	Fan *et al.,* 2016; Liu *et al.,* 2017
DBAASP	11336 peptides	http://dbaasp.org	It stores peptides data from PubMed, concerning AMPs of different routes of synthesis (ribosomal, nonribosomal and synthetic) and level of complexity (monomers, dimers and covalent-linked peptides).	Gogoladze *et al.,* 2014; Pirtskhalava *et al.,* 2015

The APD enables effective search, prediction, and design of peptides with antibacterial, antiviral, antifungal, antiparasitic, insecticidal, spermicidal, anticancer activities, chemotactic, immune modulation, or antioxidative properties and the upgraded APD makes predictions based on the database-defined parameter space and provides a list of the sequences most similar to natural AMPs. Several newly annotated activity types are unique in this database, making the APD most comprehensive in terms of activity annotation (Wang, 2015).

It is a comprehensive tool for discovery timeline, naming (nomenclature), classification, information search, statistical analysis, prediction and design of AMPs covering all six life kingdoms (bacteria, archaea, protists, fungi, plants, and animals). The peptide data stored in APD were extracted from the literature (PubMed, PDB, Google, and Swiss-Prot) manually in over a decade. The current AMP database contains over 2,950 antimicrobial peptides and proteins from living prokaryotic and eukaryotic organisms such as bacteria, archaea, protists, fungi, plants and animal host defense peptides.

The APD3 has set up criteria for peptide registration. This database currently focuses on (i) natural AMPs with (ii) a known amino acid sequence, (iii) biological activity and (iv) a size less than 100 residues although APD also includes some small yet important antimicrobial proteins >100 aa (Wang *et al.,* 2015). A few synthetic peptides of interest (~2% of the entries) are collected as separate entries and additional synthetic peptides derived from natural AMPs are annotated in their parent entries, however these derivatives are not of interest for the present research. An example of detailed information of an AMP search result from the database is shown in Table 2.2.

Table 2.2: An example of detailed information of an antimicrobial peptide search

Antimicrobial Peptide AP00016	
APD ID	AP00016
Name/Class	Aurein 2.3 (XXA, UCLL1c; frog, amphibians, animals)
Source	Southern bell frog *Litoriaaurea*or *Litoriaraniformis*, Australia
Sequence	GLFDIVKKVVGAIGSL
Length	16
Net charge	2
Hydrophobic residue%	56%
Boman Index	-1.01 kcal/mol
3D Structure	Helix
Method	NMR
SwissProt ID	SwissProt ID: P82390
Activity	anti-Gram+ & Gram-, antifungal, Cancer cells
Crucial residues	Gain anticancer activity if change I13 to F (change to P82392)
Additional info	Active against L. lactis (MIC 25 ug/ml) (provided by Chunfeng Wang). C-terminal NH2. Solution NMR in 25% TFE established a helical structure (Biophys. J 2007; 92: 2854). Found in multiple species. Updated 9/2017.
Title	The antibiotic and anticancer active aurein peptides from the australian bell frogs Litoriaaurea and Litoriaraniformis the solution structure of aurein 1.2.
Author	Rozek T, Wegener KL, Bowie JH, Olver IN, Carver JA, Wallace JC, Tyler MJ.2000
Reference	Eur J Biochem. 2000 Sep; 267(17):5330-41. PubMed.

2.2.2.2. Collection of Anti-Microbial Peptides (CAMP)

Collection of Anti-Microbial Peptides (CAMP) is a free online database that has been developed for advancement of the present understanding on antimicrobial peptides. CAMP is freely available at http://www.bicnirrh.res.in/antimicrobial (Thomas *et al.*, 2009).

It is manually curated and the sequences are divided into experimentally validated (patents and non-patents) and predicted datasets based on their reference literature. Information like

source organism, activity (minimum inhibitory concentration values), reference literature, target and non-target organisms of AMPs are captured in the database (Thomas *et al.*, 2009). Sequence and structure analysis tools have been incorporated to enhance the usefulness of the database (Waghu *et al.*, 2014).

CAMPR3 is an update to the existing CAMP database available online at www.camp3.bicnirrh.res.in. CAMPR3 presently holds 10247 sequences, 757 structures and 114 family-specific signatures of AMPs. Users can avail the sequence optimization algorithm for rational design of AMPs embedded within the database (Waghu *et al.*, 2015). Links to UniProtKB, PDB, PubMed and other databases dedicated to AMPs are also made available for the benefit of the users. The data in CAMP is organized into 17 fields namely; CAMP ID, sequence, sequence length, source, taxonomy, activity, Gram nature, target organisms, haemolytic activity, PubMed ID, protein name, protein definition, GenInfo ID, Swiss-Prot, PDB accession numbers, comments and the dataset type (experimentally validated/patents/predicted) as seen in Table 2.3. Based on their activity, peptides are classified as 'antibacterial', 'antifungal', 'antiviral' or 'antiparasitic' (Thomas *et al.*, 2009).

Table 2.3: Various fields of data in the CAMP database search

CAMPSQ983/APD00193/DRAMP03565

Title :	Hepcidin-25 / LEAP		
Source :	Homo sapiens [Human]		
Length :	25		
Activity :	Antibacterial, Antifungal		
Gram Nature :	Gram+ve, Gram-ve		
Target :	E.coli ML35p, S. epidermidis, S. aureus, C. albicans, group B Streptococcus , Neisseria cinerea , Saccharomyces cerevisiae (MIC = 50 mg/ml) , B.subtilis (MIC = 40 mg/ml)		
Validated :	Experimentally Validated		
Pfam :	PF06446 : (Hepcidin)		
InterPro :	IPR010500 : Hepcidin.		
AMP Family :	Hepcidin		
Gene Ontology :	GO ID	Ontology	Definition
	GO:0005576	Cellular Component	Extracellular region
Sequence:	DTHFPICIFCCGCCHRSKCGMCCKT		

2.2.2.3. Data Repository of Antimicrobial Peptides (DRAMP)

DRAMP (Data Repository of AntiMicrobial Peptides) is an open-access web server tool with detailed annotations of AMPs, especially detailed antimicrobial activity data and structure and sequence information including physicochemical information, patent information,

clinical information and reference information. This database is divided into three datasets: general, patent, and clinical AMPs (Fan *et al.,* 2016).

DRAMP is a manually curated database and can be accessed freely at http://dramp.cpu-bioinfor.org a comprehensive user-friendly data repository of antimicrobial peptides, which holds 17349 antimicrobial sequences, including 4571 general AMPs, 12704 patented sequences and 74 peptides in drug development. Annotations also include accession numbers cross linking to Pubmed, Swiss-prot and Protein Data Bank (PDB) (Fan *et al.,* 2016). The DRAMP ID of an entry in query or browse results gives a detailed information page. This page presents all annotations for the entry which are divided into general information, activity information, structure information, physicochemical information, comment information and literature information (Fan *et al.,* 2016).

2.2.2.4. Database of Antimicrobial Activity and Structure of Peptides (DBAASP)

The Database of Antimicrobial Activity and Structure of Peptides (DBAASP) is a freely accessible at http://dbaasp.org which contains information on AMPs of different origins (ribosomal, non-ribosomal and synthetic) and complexity (monomers, dimers and two-peptides). DBAASP is manually curated and is a depository of information on those peptides for which antimicrobial activity against particular targets has been evaluated experimentally. The database provides: 1) Full information on the chemical structure of peptides: complete information regarding post-translational and amino (N)/carboxyl (C) termini modification of amino acids. 2) Information about peptide antimicrobial activities and experimental conditions at which activities were estimated. 3) Information about peptide haemolytic/cytotoxic activities. 4) Information about the target object of the cell (Gogoladze *et al.,* 2014). In the database, information on the peptides' activities against more than 4200 different organisms (bacteria, fungi, some parasites, viruses and cancer cells) can be found.

Currently, DBAASP v.2 serves as a repository of the information necessary for the study of structure/activity relationships. DBAASP v.2 allows for the collection of statistically significant sets of peptides having experimentally validated activities against particular pathogens (Pirtskhalava *et al.*, 2015).

A new version of the database reports chemical structures and empirically-determined activities (MICs, IC50, etc.) against more than 4200 specific target microbes for more than 2000 ribosomal, 80 non-ribosomal and 5700 synthetic peptides. The main difference between DBAASP and other AMP databases is its ability to provide the data required to perform structure/activity studies (i.e. comprehensive data on chemical and 3D structures along with susceptibilities of specific pathogenic agents) (Pirtskhalava *et al.*, 2015). Full information on a peptide is presented in a peptide card seen in Table 2.4.

Table 2.4: DBAASP peptide is presented in a peptide card

DBAASP Peptide Card

ID	595
Name	Thionin
Synthesis Type	Ribosomal
Target Group	Fungus Gram- Gram+
Target Object	Lipid Bilayer
UniProt	**Peptide:** P07504
PDB	
Note	

Source and Gene

Kingdom	Subkingdom	Source	Gene	NCBI	Note
Plantae		Pyrularia pubera	THI1		

Monomer

N Terminus	Sequence	C Terminus	Length
	KSCCRNTWARNCYNVCRLPGTISREICAKKCDCKIISGTTCPSDYPK		47

Intrachain bond

Position of First Amino Acid	Position of Second Amino Acid	Bond Type	Note
3	41	DSB	
4	33	DSB	
12	31	DSB	
16	27	DSB	

Physico-Chemical Property

Normalized Hydrophobicity	Net Charge	Isoelectric Point
0.51	6	8.73

MD Structural Model

- Representative structure **View 3D (by Jmol)**
- Self-consistency
- Secondary structure
- Trajectory

Activity Against Target Species

Target Species	Activity Measure	Activity	Unit	pH	Ionic Strength mM	Salt Type	Note	Reference
Clavibacter michiganensis	EC50	0.23±0.04	μM					1
Rhizobium meliloti	EC50	>20±0.0	μM					1
Xanthomonas campestris	EC50	3.67±0.58	μM					1
Fusarium oxysporum	EC50	0.38±0.08	μM					1

References

1. Vila-Perelló M, Sánchez-Vallet A, García-Olmedo F, Molina A, Andreu D.
J Biol Chem, 2005, 280, 1661-1668, .
Structural dissection of a highly knotted peptide reveals minimal motif with antimicrobial activity
Pubmed

Latest Publications

User can select desired interval of time and in accordance with the peptide's name or UniProt ID will get additional data from PubMed.These data hasn't been yet analyzed and deposited in DBAASP. Default time interval is 30 day back from the moment of the search.

30 Search Reset

2.2.3. Biomedical text mining

The recent biomedical advances that have prevented or altered the course of many diseases are undoubtedly valued by all. Progress in biomedicine is attributable to advances in the understanding of disease mechanisms and the societal and commercial value of researching these mechanisms as well as the approaches for the prevention and cure of diseases. Biomedical text mining holds the promise of and in some cases delivers a reduction in cost and an acceleration of discovery (Simpson and Demner-Fushman, 2012). Due to the exponential growth of the biomedical literature, text mining tools have become crucial to process all available information contained in literature databases such as PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and Google Scholar (https://scholar.google.co.za/) which are accessed for free. The keyword search with PubMed offers optimal update frequency and includes online early articles; other databases rates articles by number of citations, as an index of importance. PubMed remains an optimal tool in biomedical electronic searches (Falagas *et al.*, 2008) and is the most widely used biomedical bibliographic text database (Ioannou *et al.*, 2014). Google Scholar, as for the Web in general, can help in the retrieval of even the most obscure information but, its use is marred by inadequate, less often updated, citation information (Falagas *et al.*, 2008). Information retrieval (IR) is the activity of finding documents that answer information needed and IR systems (search engines) such as Google and PubMed being designed specifically to query the databases of biomedical publications (Ananiadou *et al.*, 2006).

2.3. Computational approaches for the prediction of novel AMPs

2.3.1. Hidden Markov Models (HMM)

Algorithms of sequence alignment are the key instruments for computer-assisted studies of biopolymers e.g. a protein or DNA. Sequence alignment is the main method for comparing

biological sequences. Local alignments identify regions of similarity within long sequences that are often widely divergent overall. Local alignments are often preferable, but can be more difficult to calculate because of the additional challenge of identifying the regions of similarity (Polyanovsky *et al.*, 2011). A variety of computational algorithms have been applied to the sequence alignment problem. These also include probabilistic methods designed for large-scale database searches. Sequence alignments are stored in a wide variety of text-based file formats and most web-based tools allow a limited number of input and output formats, such as FASTA format and GenBank format (Porto *et al.*, 2017).

A Multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA or RNA to identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships between the sequences. High sequence similarity between molecules usually implies significant structural and functional similarities in an alignment. Computational algorithms are used to produce and analyse the alignments. MSAs require more sophisticated methodologies because they are more computationally complex (Mulia *et al.*, 2012).

HMMER is used for searching sequence databases for sequence homologs and for making sequence alignments. It implements methods using probabilistic models called profile hidden Markov models (profile HMMs). HMMER is designed to detect remote homologs as sensitively as possible, relying on the strength of its underlying probability models. Modelling biological sequences such as proteins and DNA sequences consisting of smaller substructures with different functions and different functional regions often displays distinct statistical properties (Yoon, 2009). HMMs have been shown to be very effective in representing biological sequences, and have become increasingly popular in computational molecular biology with many state-of-the-art sequence analysis algorithms being built on HMMs (Yoon, 2009).

The HMMER software is designed to run on UNIX platforms Linux and Apple Macintosh. All the HMM profiles in this study were constructed using Ubuntu 12.04 LTS operating system, which is based on the Linux kernel. Previous versions of HMMER have largely only been available as computationally intensive UNIX command line applications requiring local installation and local computing resources (Finn *et al.*, 2011).

2.4. Aims

The aim of this chapter was to construct a sensitive and specific probabilistic model with experimental validated anti-*N. meningitidis* AMPs as input for identification of putative anti-*N. meningitidis* AMPs from various proteome sequences. With the objectives of this chapter being as follows:

- Extract and generate an anti-*N. meningitidis* AMP list by collecting all peptides within various databases and verify that they are experimentally validated as having activity against *N. meningitidis* through data mining. Cross reference between databases to eliminate duplicate entries within the final list.
- Extract the amino acid sequences for AMPs within the final list
- Construct Hidden Markov Models from the amino acid sequences.
- Test the constructed models as to optimize them.
- Scan the optimized models against various genome sequences
- Identify potentially novel anti-*N. meningitidis* AMPs

Section 2.5.1 presents the retrieval of experimentally validated anti- *N. meningitidis* AMPs from various databases and the elimination of duplicates. Section 2.5.2 will present the procedure for constructing HMMER models and finally, section 2.5.3 gives the performance measures of the models created using HMMER and section 2.5.4 scanning of the genome sequences to identify novel *N. meningitidis* AMPs.

2.5. Materials and methods

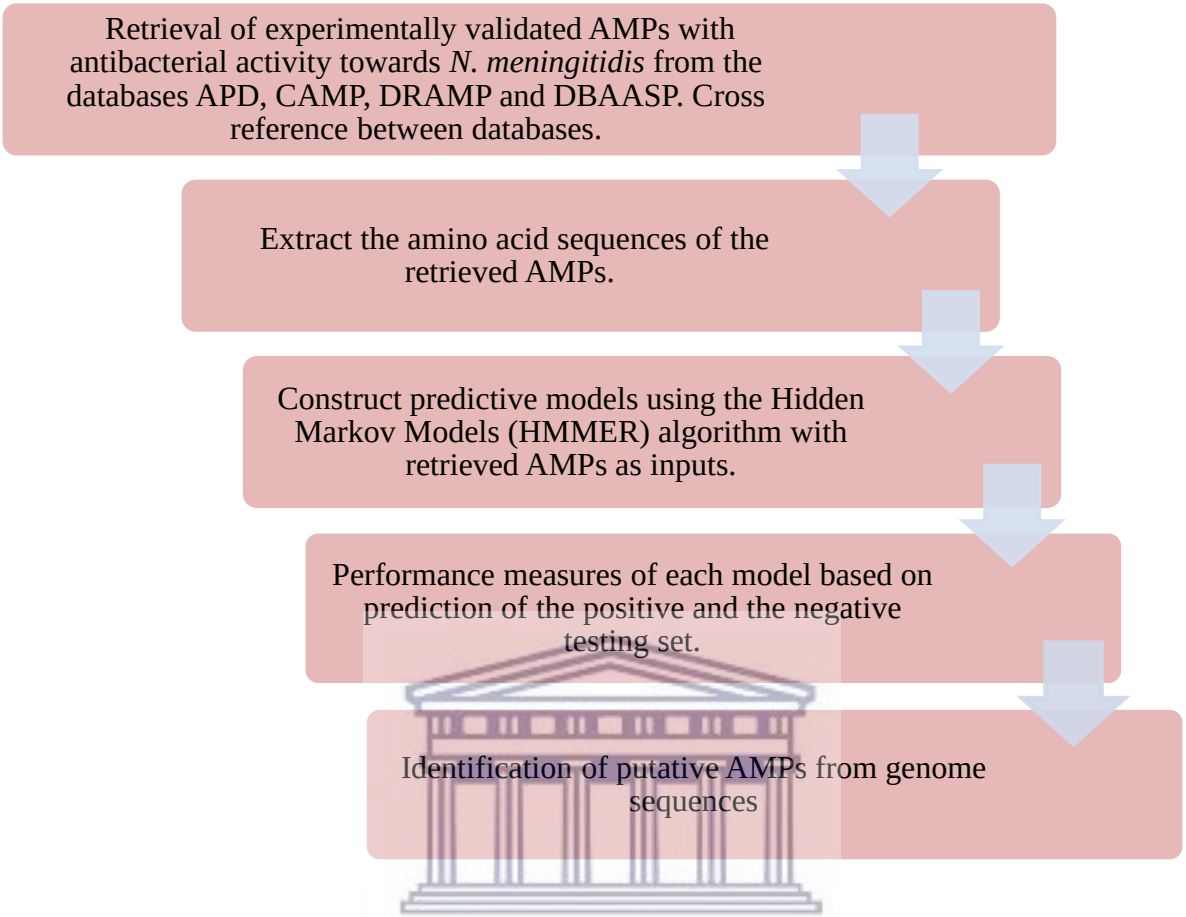


Figure 2.2: Flowchart describing the outline of methodology to retrieve novel antimicrobial peptides (AMPs)

2.5.1. Data mining: Experimentally Validated Antimicrobial Peptides (AMPs) Data Assessment

2.5.1.1. Data retrieval using the APD database

To retrieve the experimentally validated AMPs, the APD database was accessed at (http://aps.unmc.edu/AP/) and from the homepage the activity 'Antibacterial peptides' was selected. This resulted in a new page displaying the peptides requested. Data mining and literature mining were done simultaneously as each peptide was scanned for activity against the genus *Neisseria* but specifically against *N. meningitidis*. The criteria used for the database

search were only natural (not synthetic), experimentally validated AMPs. The amino acid sequences of these AMPs were extracted and saved in the FASTA format as a text document.

2.5.1.2. Extracting anti–*N. meningitidis* AMPs from CAMP

The extraction of the available AMPs with antibacterial activity against *N. meningitidis* from the CAMP database http://www.bicnirrhx.res.in/antimicrobial/index.php was done by accessing the homepage and "AMP Sequences" selected. The criteria used and selected were activity 'Antibacterial' and validation 'Experimentally Validated (excluding Patents)'. Excluded in the search were synthetic constructs as the source, predictive AMPs and gram positive bacteria. Again data and literature mining were done simultaneously. Upon completion of searching the database, antibacterial peptides were saved as an excel spreadsheet for cross referencing.

2.5.1.3. Cross referencing the retrieval of AMPs using DRAMP

DRAMP database was accessed as a confirmation of results obtained from APD and CAMP and as a cross reference. 'Search' was selected on the homepage followed by 'Advanced search'. The selection was made of "Antibacterial" as the choice for biological activity to limit the search and the 'Submit' button was selected.

The AMP list was analysed by searching each AMP with reference information and with the same criteria as for the APD and CAMP database searches: 'experimentally validated' and 'natural'. A list of AMPs was created and further refined by removal of duplication.

2.5.1.4. Further AMP extraction and identification of anti-*N. meningitidis* within the DBAASP database

The DBAASP database was accessed on http://dbaasp.org and on the homepage 'Search' was selected, resulting in a new input window. The default settings were used except for the target species selected as '*Neisseria Meningitidis*' and 'Search' button selected. Another selection

was made which included the target species as '*Neisseria spp*'. All experimentally evaluated AMPs were selected and synthetic AMPs excluded under 'synthesis type' within the DBAASP peptide card. AMPs were then cross reference between the four databases. See Table 2.5 as an example.

2.5.1.5. Literature mining of the AMPs

The databases used for literature mining for selected AMPs with calculated minimum inhibitory concentration (MIC) as experimental evidence against *N. meningitidis* were: PubMed and Google Scholar. All relevant literature, abstracts and journal articles were searched for information linking each AMP to the criteria selected as having antimicrobial activity to either the species *Neisseria* or more specifically *N. meningitidis*.

2.5.1.6. Removal of duplicates and generation of final AMP list

The final elimination of duplicates of the experimentally validated AMPs peptides was done between databases with cross referencing. The elimination was made on the basis of the peptide name and done manually since the numbers of AMPs were minimal. Due to the small amount of AMPs retrieved for the bacteria *N. meningitidis*, AMPs with antibacterial activity against the genus *Neisseria* were also included in the final AMP list. The retrieved AMPs were recorded in FASTA format.

2.5.2. Construction of Hidden Markov Models (HMM)

2.5.2.1. Construction of the Training and testing data sets

The Hidden Markov Models algorithm version 2.3.2 was used to create HMMs profiles by utilizing AMP training sets as the input data and the testing set data, to strengthen the predictive ability of the profiles.

The final list of the experimentally validated AMPs were too small to create a robust predictive model and therefore three text files in FASTA format named All AMPs, *N. gonorrheoae* (*Ngon*) and *N. meningitidis* (*Nmen)* respectively were created, hence three different models were created instead of the anticipated one model.

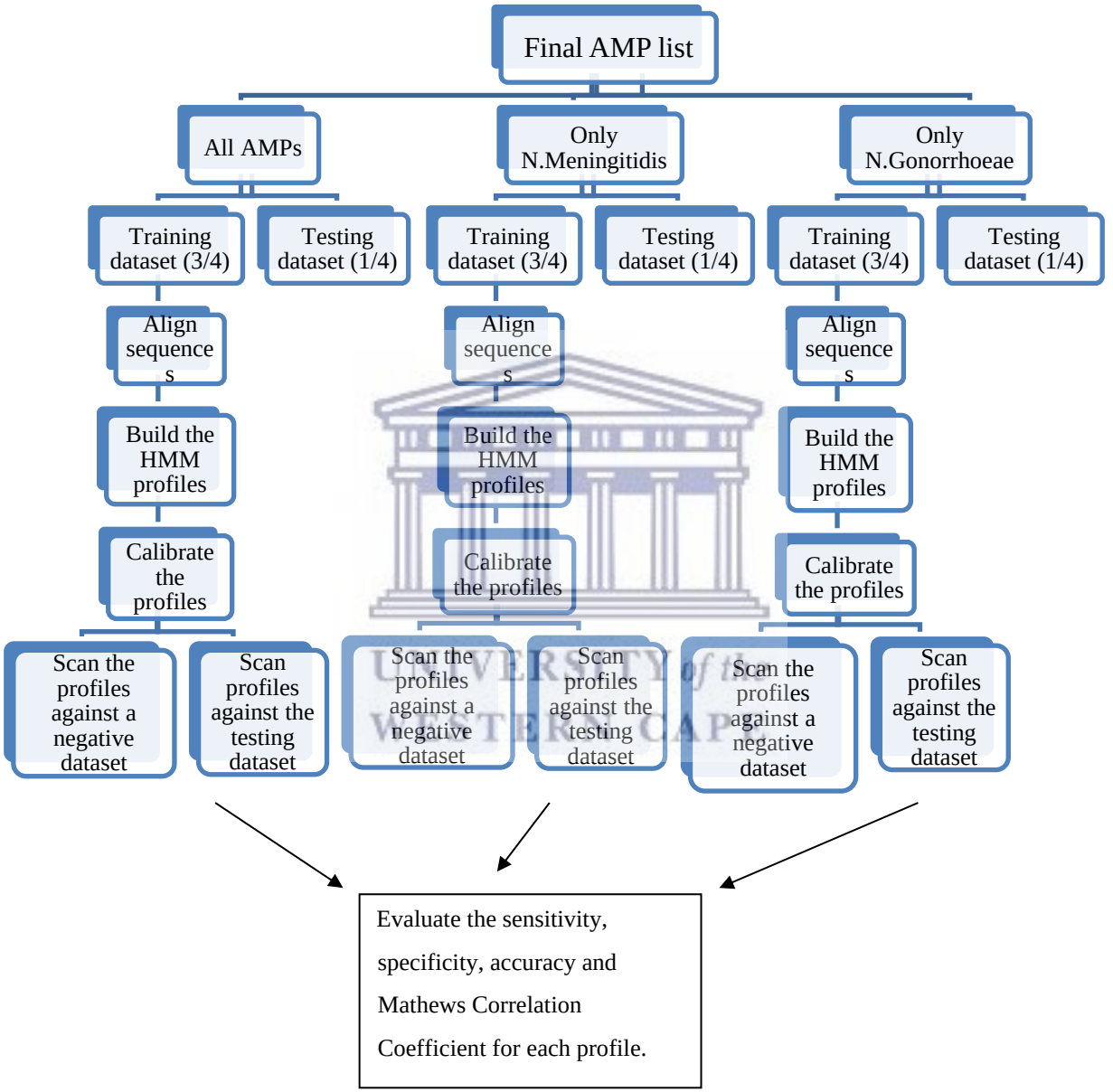
Each AMP data set was then divided into two portions: three-quarters of each data set was utilized as the training set whilst one-quarter was used as the testing set as seen in Figure 2.3.

2.5.2.2. Sequence alignment of training set

The HMMER software was installed and using ClustalW (Larkin *et al.*, 2007), an alignment was done for the input sequences of the training set for All AMPs, *Ngon* and *Nmen* using the command line:

```
Clustalw -align -output=gcg -case=upper -seqnos=off -outorder=aligned -infile=
filename.fasta
```


Figure 2.3: Outlined the method to build profiles using the profile HMM algorithm



The command line of ClustalW means “to do an alignment of the sequences which are in the upper case found in the input filename e.g. “AllAMPstraining” with the FASTA format, using ClustalW as multiple alignment tool and GCG Postscript output for graphical printing”.

The result was saved as msf (gcg) format (filename.msf). This aligned sequence was used for the next step.

2.5.2.3. Create a HMM profile from the aligned sequences

To search for additional remote homologues and using the hmmbuild module common motifs within the model was achieved by inputting the hmmbuild command line, which builds a new HMM profile from the multiple sequence alignment file using the command line:

```
hmmbuildfile.hmmfile.msf
```

The ‘build profile’ was saved in hmm format (filename.hmm).

2.5.2.4. Calibrate the profile HMM to enhance sensitivity

hmmcalibrate - this step calibrates the HMM search statistics. This command line helps to improve the profile sensitivity.

```
hmmcalibratefilename.hmm
```

Models were calibrated using the training dataset in order to increase sensitivity. The resulting profile ‘training.hmm’, a calibrated model was employed in evaluating the performance of the profile by scanning it against anti-*N. meningitidis* AMPs in the testing sets designed.

2.5.3. Model Testing

A profile HMM searches against a target sequence database, with the profile HMM being built from a query multiple sequence alignment. To measure the performance of the constructed profiles of the training datasets, the testing set of each dataset was used to query the profile of that dataset.

This step helps to predict if the testing set of a particular dataset really belongs to the profile for that dataset i.e. if the sequences of the testing set belong to the constructed models which were built based on primary sequences of the training set e.g. positive dataset.

The profile also confirms that the testing and the training sets have anti-*N. meningitidis* activities since both sets (training and testing) were derived from the list of experimentally validated anti-*N. meningitidis* AMPs.

Following the profile calibration, the hmmsearch step was carried out, a step which searches and queries a particular peptide or protein list. Testing AMPs set (one-quarter of the retrieved AMPs) was utilized as the positive dataset hence, the testing AMPs set was queried against the constructed profile. The search was achieved using the command line:

The E-value threshold (list of ranked top hits sorted by E-value, most significant hit first) was set to 1% or 0.01 as indicated on the command line, thus to increase sensitivity and the scanning was achieved using a new command line:

```
hmmsearch -E 1e-2 filename.hmm filename(test set).fasta >
resultname.txt
```

The scores and E-values here reflect the confidence that this query sequence contains one or more domains belonging to *Neisseria.*

Negative set evaluation

The profiles were also tested on a negative dataset consisting of 780 peptide sequences which are non-anti-*Neisseria* peptides. The negative set are the sequences with no activity against the bacteria *Neisseria* whilst the testing set has activity against *Neisseria* and thus is the positive set.

HMMER reads all major database formats and does not need any special database indexing thus the negative set was downloaded in FASTA format.

Using the command line:

```
hmmsearch -E 1e-2 filename.hmm negativeset.fasta> resultname.txt
```

This step helps to identify the number of true positive, false negative, true negative and false negative antimicrobial peptides, which enabled the performance calculation of the constructed models. Testing the robustness of each HMMER profiles by scanning it against an anti-*N. meningitidis* AMP testing set and evaluates its performance and the algorithms ability to identify and highly discriminate putative AMPs from non-anti-*N. meningitidis* AMPs using the negative set.

2.5.4. Performance measures of each profile based on prediction of the positive and the negative testing set

By using sensitivity, specificity, accuracy and Mathew Correlation Coefficient as the parameters to measure the statistical performance evaluation of the created model, the performance of the profile was calculated. Models were measured for the evaluation of the quality of the prediction and were calculated to assess the strength of the constructed models. The reliability of these methods is evaluated by several parameters seen in Table 2.5.

Table 2.5: Definitions of the metrics used to formulate the evaluation of the quality of the prediction

Metric	Description	Definition
TP	True Positive	represents correctly predicted anti-*N. Meningitidis* AMPs
TN	True Negative	correctly predicted negative AMPs
FN	False Negative	number of AMPs incorrectly predicted as non anti-*N. Meningitidis* AMPs
FP	False Positive	number of non anti-*N. Meningitidis* AMPs incorrectly predicted as AMPs
MCC	Mathew's correlation coefficient	indicates completely random or perfect prediction

The statistical measures were calculated to evaluate the strength and the performance of the created model using the following mathematical formulas;

Sensitivity is the percentage of anti-*N. meningitidis* AMPs (testing set) correctly predicted as anti-*N. meningitidis* AMPs (Positive)

$$Sensitivity = \left(\frac{TP}{TP + FN}\right) \times 100 \qquad (1)$$

Specificity is the percentage of non-anti-*N. meningitidis* AMPs (negative set) correctly predicted as non-anti-*N. meningitidis* AMPs (negative). The specificity is defined as:

$$Specificity = \left(\frac{TN}{TN + FP}\right) \times 100 \qquad (2)$$

Accuracy is the percentage of correctly predicted anti-*N. meningitidis* and non-anti-*N. meningitidis* AMPs. The accuracy is defined as:

$$Accuracy = \left(\frac{TP + TN}{TP + FP + TN + FN}\right) \times 100 \qquad (3)$$

Mathew's correlation coefficient (MCC) is a measure of both sensitivity and specificity. It is worthy to note that MCC = 0 indicates completely random prediction, while MCC = 1 indicates perfect prediction. It is defined as:

$$MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}} \qquad (4)$$

Through these sets the algorithm is trained and then, tested against the negative set. From the results against the negative set, the true positives, negative and false positives and negatives are estimated and then the parameters (*e.g.*, accuracy) are calculated.

2.5.5. Scanning of proteomes using profile HMM based on anti-*N. Meningitidis* AMPs

A list of all proteome sequences were retrieved from the ENSEMBL server (http://www.ensembl.org/index.html) and UNIPROT database (http://www.uniprot.org/) of an amount of 959 genomes in FASTA format. The protein genomes were retrieved from insects, microbes, plants, fish, animals and amphibians.

In order to identify the putative anti- *N. Meningitidis* AMPs, the created profile/s was queried against all the proteome sequences with an E-value cut-off of 0.01. This was achieved by utilizing the hmmsearch module of HMMER and the command line used for scanning the proteome sequences:

```
hmmsearch -E 1e-2 filename.hmm filename (genomes).fasta >
resultname.txt
```


After performance calculation and scanning of various genome sequences databases duplicate sequences were removed to have a final list of putative AMPs.

2.6. Results and Discussion

2.6.1. Mining of biologic datasets

The methodology used in this research work was in line with the work of Tincho *et al.*, 2016, "In-*Silico Identification and Molecular Validation of Putative Antimicrobial Peptides for HIV Therapy*". The Antimicrobial Peptide Database (APD) was used to extract all natural, experimentally validated antibacterial peptides having activity against the *N. meningitidis*

bacteria. The APD created a set of criteria to collect peptides with the aim of guaranteeing curation quality (Liu *et al.,* 2017). The initial search of the database resulted in 2415 antibacterial peptides which were experimentally validated, inferred and synthetic antibacterial peptides. Search of the latest version of APD (APD3) resulted in 2434 peptides identified. The search in CAMP database resulted in 2157 antibacterial peptides and 2955 antibacterial peptides in the DRAMP database search. A list of AMPs was created and further refined by removal of duplication resulting in the same AMPs generated from APD and/or CAMP.

Data and literature mining revealed after cross referencing (example seen in Table 2.7) between the databases APD, CAMP, DRAMP and DBAASP, 20 AMPs experimentally validated as anti-*N. meningitidis* peptides. AMPs were considered experimentally validated based on molecular testing against certain microbes/bacteria with MIC values indicated for these AMPs. Peptides were extracted with activity against the bacteria *N. meningitidis* as well as other *Neisseria* genus e.g. *N. cinerea*, whilst other peptides have activity against only the two pathogens in the *Neisseria* family i.e. *N. gonorrhoeae* and *N. meningitidis*. See Table 2.6 for the breakdown of AMPs extracted. Due to the small amount of AMPs extracted, the *Neisseria* genus was included for the creation of prediction models. The origin or source of the peptides was considered only if it was not synthetic or inferred i.e. only natural AMPs.

It has to be noted that since there was only a small amount of AMPs retrieved and after all the four databases were exhausted in the search, various other database were also searched such as; Cybase (Mulvenna *et al.,* 2006, Wang *et al.,* 2007), Bactibase (Hammami *et al.,* 2010) and DAMPD (Sundararajan *et al.,* 2011). The Cybase database resulted in only 12 AMPs with antibacterial activity but none with activity against *N. meningitidis*. Only one AMP in the Bactibase database showed activity to the genus *Neisseria* but resulted in being a

duplicate found in the CAMP database. The DAMPD database (Sundararajan *et al.,* 2011) was not accessible (URL not found on server) at the time of the search.

Table 2.6: Number of AMPs extracted after database searches

Number of AMPs retrieved	Neisseria family type
11	N. gonorrhoea only
1	N. meningitidis only
2	Both pathogens N. meningitidis and N. gonorrhoea
2	N.Cinerea only
3	N. meningitidis with other family
1	Neisseria species only

All databases, share some sequences with at least one other database, with the largest mutual overlap being single peptide commonalities. Moreover, there is some degree of redundancy in each database (Aguilera-Mendoza *et al.,* 2015) and this was evident in the four databases searched for AMPs with activity against *Neisseria* based on the results obtained from data mining.

2.6.2. Creation of training and testing sets for HMMs

The sequences of the All AMPs, *Ngon* and *Nmen* files were used to create the testing and training sets for profile creation using HMM.

The sequences of the anti-*Neisseria* AMPs of the three files were randomly divided into two portions: three-quarters of each data set was utilized as the training set, whilst one-quarter

was used as the testing set. The model created from the abovementioned partitioning failed to perform well after evaluation. To create a model that was more robust, sequences within each file, All AMPs, *Ngon* and *Nmen*, were randomly divided into three quarter training and one quarter testing set resulting in six datasets for each file to be used for model creation (Table 2.8).

2.6.3. Evaluate the algorithm performance measurements of the six model creations

After testing all six models, performance measurements were calculated. The following measures were used; sensitivity, specificity, accuracy and MCC as seen in Table 2.8. The accuracy represents how well a method can predict in a range from 0 to 100 percent. Larger values equate to better classification performance, given that the number of positive and negative examples are comparable in size. MCC is a summary statistic to evaluate binary classification performance and is more stringent compared to accuracy, as it weighs the impact of false predictions more heavily. Values can range from 0 to 1. Larger values equate to better classification performance (Veltri, 2015). Performance analysis to identify the most robust model created for each file is reported in Table 2.8 in terms of sensitivity, specificity, accuracy and MCC.

Table 2.7: Cross referencing between the various databases with similar peptides

AMP extraction from various databases				
	APD	**CAMP**	**DRAMP**	**DBAASP**
ID	AP00050	CAMPSQ38	DRAMP01097	529
Peptide name	Bombinin-like peptide 1 (XXA, BLP-1, UCLL1c; toad, amphibians, animals)	Bombinin-like peptides 1	Bombinin-like peptide 1 (toads, amphibians, animals)	Bombinin-like peptide 1, BLP-1
Source	Oriental fire-bellied toad/frog, Bombina orientalis, Asia	Bombina orientalis [Oriental fire-bellied toad]	Bombina orientalis (Oriental fire-bellied toad)	Bombina orientalis
Sequence	GIGASILSAGKSALKGLAKGLAEHFAN	GIGASILSAGKSALKGLAKGLAEHFAN	GIGASILSAGKSALKGLAKGLAEHFAN	GIGASILSAGKSALKGLAKGLAEHFAN
Sequence length	27	27	27	27
Pubmed ID	1744108	1744108	1744108	1744108
Ref	J. Biol. Chem. 1991; 266: 23103-23111. PubMed.	J Biol Chem. 1991 Dec 5;266(34):23103-23111.	J Biol Chem. 1991 Dec 5;266(34):23103-23111.	Gibson BW, Tang DZ, Mandrell R, Kelly M, Spindel ER. J Biol Chem, 1991, 266, 23103-23111.
Validated Experimentally	Yes	Yes	Yes	Yes
Target	Active against N. meningitidis strains 118V, 15240, 126E, N. gonorrhoeae, N. lactamica, and N. cinerea (82-100% at 20 ug/ml).	82 % inhibition Neisseria meningitidis 118V C (MIC = 20 microg/ml),98 % inhibition Neisseria meningitidis 15240 IAI(MIC = 20 microg/ml), 85% inhibition Neisseria meningitidis 126E IC1(MIC = 20 microg/ml),98% inhibition Neisseria gonorrhoeae F62(MIC	82 % inhibition Neisseria meningitidis 118V C (MIC=20 µg/ml), 98 % inhibition Neisseria meningitidis 15240 IAI(MIC=20 µg/ml), 85% inhibition Neisseria meningitidis 126E IC1(MIC=20 µg/ml), 98% inhibition Neisseria gonorrhoeae F62 (MIC=20 µg/ml), 96% inhibition Neisseria lactamica 15323(MIC=20 µg/ml), 100% inhibition Neisseria lactamica 15215A(MIC=20 µg/ml), 98% inhibition Neisseria cinerea 15461 (MIC=20 µg/ml),	82 % inhibition Neisseria meningitidis 118V C (MIC=20 µg/ml), 98 % inhibition Neisseria meningitidis 15240 IAI(MIC=20 µg/ml), 85% inhibition Neisseria meningitidis 126E IC1(MIC=20 µg/ml), 98% inhibition Neisseria gonorrhoeae F62 (MIC=20 µg/ml), 96% inhibition Neisseria lactamica 15323(MIC=20 µg/ml), 100% inhibition Neisseria lactamica 15215A(MIC=20 µg/ml), 98% inhibition Neisseria cinerea 15461 (MIC=20 µg/ml),
Title	Bombinin-like peptides with antimicrobial activity from skin secretions of the Asian toad, Bombina orientalis.	Bombinin-like peptides with antimicrobial activity from skin secretions of the Asian toad, Bombina orientalis.	Bombinin-like peptides with antimicrobial activity from skin secretions of the Asian toad, Bombina orientalis.	Bombinin-like peptides with antimicrobial activity from skin secretions of the Asian toad, Bombina orientalis.

As seen in Table 2.8, yellow highlighted model 3 was the best model showing the sensitivity as 40%, 25% and 50% for All AMPs, All *Ngon* and All *Nmen*, respectively. This result measures the proportion of actual positives that are correctly identified. Specificity was calculated as 100% for All AMPs, *Ngon* and *Nmen* models created. The accuracies of the models were calculated as 99.60%, 99.62% and 99.87% for All AMPs, All *Ngon* and All *Nmen*, respectively. This proves that the models have more than 95% confidence to predict a peptide as a putative anti-*N. meningitidis* AMP. The MCC values calculated for model 3 for all data subsets indicates that the algorithm is performing similarly to perfect prediction as the values are closer to 1. Thus from the performance measures for all the models created, model 3 for each dataset scored the best across all the performance measures calculated and will subsequently be used for genome scanning.

Table 2.8: Performance measurements generated for each model created by HMMER profile.

Model	Sensitivity	Specificity	Accuracy	MCC
All AMPs	20%	100%	98.50%	0.44
All Ngon	25%	100%	98.90%	0.49
All Nmen	50%	100%	99.60%	0.71
All AMPs 2	0%	100%	99.36%	0
All Ngon 2	0%	100%	99.49%	0
All Nmen 2	0%	100%	99.74%	0
All AMPs 3	40%	100%	99.60%	0.63
All Ngon 3	25%	100%	99.62%	0.499
All Nmen 3	50%	100%	99.87%	0.71
All AMPs 4	0%	99.87%	99.24%	0.0029
All Ngon 4	0%	100%	99.49%	0
All Nmen 4	50%	100%	99.87%	0.71
All AMPs 5	0%	100%	99.36%	0
All Ngon 5	0%	100%	99.49%	0
All Nmen 5	0%	100%	99.74%	0
All AMPs 6	40%	100%	99.62%	0.63
All Ngon 6	0%	100%	99.49%	0
All Nmen 6	50%	100%	99.87%	0.71

Genome scanning

The chosen model 3 were scanned against protein genomes from insects, microbes, plants, fish, animals and amphibians, retrieved from the ENSEMBL server and UniProt database. The total number of genomes scanned was 959, in order to search for novel anti-*Neisseria* AMPs by identifying peptides with similar motifs and properties as the created model. An example of a scanning result is seen in Figure 2.4:

```
hmmsearch - search a sequence database with a profile HMM
HMMER 2.3.2 (Oct 2003)
Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
Freely distributed under the GNU General Public License (GPL)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 N.gonorrhoeaetraining3.hmm [N.gonorrhoeaetraining3]
Sequence database:        Gallus_gallus.fasta
per-sequence score cutoff:  [none]
per-domain score cutoff:    [none]
per-sequence Eval cutoff:   <= 0.01
per-domain Eval cutoff:     [none]
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Query HMM:   N.gonorrhoeaetraining3
Accession:   [none]
Description: [none]
  [HMM has been calibrated; E-values are empirical estimates]

Scores for complete sequences (score includes all domains):
Sequence           Description                          Score    E-value  N
--------           -----------                          -----    ------- ---
GENSCAN00000033381 pep:genscan chromosome:WASHUC2:1:1   26.6    0.00039   4

Parsed for domains:
Sequence           Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
--------           ------- ----- -----    ----- -----       -----  -------
GENSCAN00000033381   2/4     643   668 ..     1    27 []     6.9       30
GENSCAN00000033381   3/4     980  1005 ..     1    27 []     6.9       30
GENSCAN00000033381   4/4    1489  1514 ..     1    27 []     6.9       30
GENSCAN00000033381   1/4     177   202 ..     1    27 []     5.9       44
```

Figure 2.4: HMMER results page Classification results of a query sequence using HMM profile.

As seen in Figure 2.4, the grey highlighted section explains the program ran using one of the models created, in this example its model 3 of the Ngon dataset scanned against the proteome of *Gallus gallus* as well as indicated the cutoff's used.

The turquoise section shows the top sequence identified (highest similarity to the model created), ranking the top hits (sorted by E-value, most significant hit first). It states the name of the target sequence followed by the description line for the sequence.

After proteome scanning, sequences were extracted and resulted in All AMPs, All *Ngon* and All *Nmen* with 139, 415 and 112 sequences, respectively. Removal of duplicates showed All AMPs with 72 sequences, All *Ngon* with 131 and All *Nmen* with 35 sequences and this included unique/single and multiple domains.

Unique/single domains and multiple domains were extracted and single domains were considered, as the single domains had complete sequence with activity against *Neisseria* and met the cut-off E-value of 0.01, whereas with multiple domains, the entire sequence/protein will not have activity against the chosen bacteria (only certain domains/parts). The E-values are much higher than the requested cut-off set at 0.01 for sequences with multi domains. The smallest E-value is indicative of a peptide that is most likely to be an AMP; this E-value gives more confidence about the probability that the predicted peptide is to be a true anti-*N. meningitidis* AMP. A final list of nine AMPs were identified and named YYNN1 – YYNN9 and the scores (bits) with an E-value for each AMP provided (Appendix A). The smallest E-value was observed for YYNN8 and then YYNN9 indicating that these peptides are most likely to have the best activity against *N. meningitidis* and potentially the highest binding affinity for the receptors of the bacterium.

These nine AMPs were used for further *in silico* interaction studies with *N. meningitidis* receptors (to be identified in Chapter 6) for the continuation of the study.

2.7. Summary

The aim of this chapter was to construct a sensitive and specific probabilistic model with experimental validated anti-*N. meningitidis* AMPs as input for identification of putative anti-

N. meningitidis AMPs from various genome sequences. This was achieved by extracting and generating an AMP list by collecting all peptides within the databases; APD, CAMP, DBAASP and DRAMP and verifying that they are natural, experimentally validated as having activity against *N. meningitidis* through data and literature mining. Cross-referencing between databases was done to eliminate duplicate entries within the final list with the final list containing 20 AMPs. The amino acid sequences of the 20 AMPs were extracted and used to construct Hidden Markov Models within HMMER and then tested using several performance measures as to optimize them.

Fjell *et al.*, 2008 published a study using HMMs to screen for AMPs in the bovine genome, which led to the discovery of a previously unknown AMP, this study highlighting an approach to design AMPs with great success. The use of the HMMER algorithm is deemed an appropriate tool, which enables a more sophisticated search for novel peptides through proteome sequence scanning (Tincho *et al.*, 2016).

The optimized models were scanned against various genome sequences retrieved from the ENSEMBL server and UniProt database and identified a number of peptide sequences that can potentially be considered anti-*N. meningitidis* AMPs. A final list of nine AMPs were identified and named YYNN1 – YYNN9 with the smallest E-values observed for YYNN8 followed by YYNN9.

Before this era, methods for *de novo* AMP discovery relied on long-standing bioinformatics methods, including sequence alignment and homology modelling for prediction of biological activity. Now, the convergence of innovations in machine learning models, the presence of modern computational tools and the availability of high-quality datasets has enabled the machine learning-aided design of AMP candidates (Lee *et al.*, 2017).

Chapter 3

Identification of *N. meningitidis* receptors and their associated pathways and secretion mechanisms to serve as targets for the putative AMPs

3.1. Introduction

Current dogma dictates that the antimicrobial activity of peptides is mediated through interaction of AMPs with target cell membranes and subsequent membrane disruption (Kulagina *et al.,* 2007). For bacteria with a largely extracellular lifestyle, such as *N. meningitidis*, their target receptors may be expected to be expressed at the surfaces of cells (Sa E Cunha *et al.,* 2009). Manifold interactions with the host are typical of invasive MD and among these, pathogen–endothelium interactions are crucial in the development of invasive MD (Simonis and Schubert-Unkmeir, 2016).

Years of research in the areas of biochemistry (Epand *et al.,* 2008) and molecular dynamics (Fjell *et al.,* 2012) suggest AMP models need to take into consideration the lipid membrane composition of a target bacteria if AMP-membrane interactions is to be achieved. Adherence can be defined as a phenomenon resulting from the interaction between two surfaces, with the participation of physical, chemical and biological factors, with contact between the bacterium and the cell being necessary for adherence to take place (Uberos *et al.,* 2015). The first contact between the bacterium and host cells involves the process of adhesion, which can depend on the interaction of specific bacterial surface molecules (Hung and Christodoulides, 2013).

The outer membrane, an essential organelle of Gram negative bacteria, is composed of four major components: lipopolysaccharide, phospholipids, β-barrel proteins and lipoproteins. The mechanisms underlying the transport of these components to outer membranes are under extensive examination (Tokuda, 2009).

For many bacterial pathogens, a multitude of methods are used to invade mammalian hosts, damage tissue sites and thwarting of the immune system from responding with one essential component of these strategies, being the secretion of proteins. Secreted proteins can play many roles in promoting bacterial virulence, from enhancing attachment to eukaryotic cells, to scavenging resources in an environmental niche, to directly intoxicating target cells and disrupting their functions. Many pathogens use dedicated protein secretion systems (Green and Mecsas, 2016). To interact with its environment, *N. meningitidis* transports many proteins across the outer membrane to the bacterial cell surface and into the extracellular medium for which it deploys the common and well-characterized autotransporter, two-partner and type I secretion mechanisms, as well as a recently discovered pathway for the surface exposure of lipoproteins (Tommassen and Arenas, 2017).

The quest for novel vaccine antigens has enormously stimulated research into cell-surface-exposed and secreted proteins in *N. meningitidis* and has led to the discovery of new transport mechanisms and machineries of general (micro) biological significance, such as the BAM (β-Barrel Assembly Machinery) and SLAM (Surface Lipoprotein Assembly Modulator) (Tommassen and Arenas, 2017). In most Gram-negative bacteria, cell surface exposed, particularly the integral OMPs (Outer Membrane Proteins) of the BAM systems are attractive targets for the development of new antimicrobials (Urfer *et al.*, 2016).

The surface of many Gram-negative bacteria contains lipidated protein molecules referred to as surface lipoproteins (Hooda *et al.*, 2017). Gram negative bacteria utilise many systems to translocate proteins into (insertion) or through (translocation) one or both membranes (Holland, 2010), as explained in section 3.2.

3.2. Pathway discovery for the Transport of Outer Membrane Components

Secretion pathways are a collection of specialized, structurally distinct pathways that deliver proteins into the extracellular space (or directly into other bacteria or eukaryotic cells). In some pathways proteins are secreted across both membranes in one step (e.g., type III and IV pathways), while in other pathways proteins are secreted sequentially across the two cell membranes (e.g., type II and V pathways). The structure and function of the proteins secreted by each pathway vary considerably, but in general the proteins play important roles in nutrient acquisition, survival in specific environments, or virulence (Wilson and Bernstein, 2016). These systems are mostly responsible for the translocation and secretion of proteins across the outer membrane of Gram-negative bacteria (sometimes employing the Sec or Tat machinery for transport across the cytoplasmic membrane). In bacteria, secretory proteins cross the cytoplasmic membrane either via the general secretion pathway (Sec-pathway) or the twin arginine translocation (Tat-pathway) (Natale *et al.,* 2008).

3.2.1. β-Barrel Assembly Machinery (BAM)

Barrel assembly machine (Bam) complex is a hetero-oligomeric complex that catalyses the integration of β-barrel proteins (proteins that fold into a closed cylindrical structure) into the bacterial outer membrane (Wilson and Bernstein, 2016). The essential OM protein BamA (originally termed Omp85 in *N. meningitidis* has been shown to be crucial for autotransporter (AT) biogenesis. BamA is the core of the essential BAM complex responsible for integration of outer membrane proteins (Jacob-Dubuisson *et al.,* 2013).

The β-barrel assembly complex, consisting of BamA, catalyses the insertion of virtually all β-barrel OM proteins (Knowles *et al.,* 2009). The Bam complex recognizes a C-terminal motif in β-barrel proteins. BamA consists of an N-terminal periplasmic domain that contains five

polypeptide transport associated (POTRA) domains and a C-terminal 16-stranded β-barrel domain (Noinaj *et al.*, 2017).

The BAM complex consists of five components: the essential core component BamA, which is an OMP itself, and four accessory lipoproteins termed BamB, BamC, BamD and BamE, each containing an N-terminal post-translational lipid modification that anchors them to the inner leaflet of the outer membrane. The BAM complex can also partner with the translocation and assembly module (TAM) complex to mediate the biogenesis of some autotransporters (Noinaj *et al.*, 2017).

3.2.2. Translocation and Assembly Module (TAM)

The TAM complex is involved in the secretion of some AT proteins with a β-helical passenger domain. Based on these findings, it has been proposed that the BAM complex assembles the β-barrel and the TAM complex assists in β-helix secretion (Selkrig *et al.*, 2012). TamA and TamB form the Translocation and Assembly Module (TAM) complex involved in the transport and assembly of a subset of ATs (Jacob-Dubuisson *et al.*, 2013).

3.2.3. Twin Arginine Translocation (TAT)

The TAT system has mostly been implicated in the secretion of folded and/ or cofactor containing proteins (Natale *et al.*, 2008; Patel *et al.*, 2014). This pathway is critical because not all proteins can be secreted in their unfolded state. This pathway consists of 2–3 components (TatA, TatB, and TatC). In Gram-negative bacteria, TatB and TatC bind a specific N-terminal signal peptide containing a “twin” arginine motif on folded Tat secretion substrates. TatB and TatC then recruit TatA to the cytoplasmic membrane, where it forms a channel. Folded proteins are then translocated across the channel and into the periplasm (Green and Mecsas, 2016).

3.2.4. Secretion (SEC) Pathway

The Sec system is involved in both the secretion of unfolded proteins across the cytoplasmic membrane and the insertion of membrane proteins into the cytoplasmic membrane (Natale *et al.,* 2008). The Sec system constitutes the general export pathway for periplasmic and outer-membrane proteins (OMPs) (Grijpstra *et al.,* 2013). The Sec system is also the major system for the insertion of helical membrane proteins into the IM (Inner Membrane) (Natale *et al.,* 2008). Most bacterial secretory proteins pass across the cytoplasmic membrane via the translocase, which consists of a protein-conducting channel SecYEG and an ATP-dependent motor protein SecA. The ancillary SecDF membrane protein complex promotes the final stages of translocation. SecA is a motor protein that uses ATP as energy source and threads the unfolded polypeptide through the channel. The adjoining SecDF complex is involved in later stages of protein translocation and presumably pulls translocating proteins from the channel at the periplasmic side of the membrane (a Nijeholt and Driessen, 2012).

A removable signal sequence recognized by the SecB protein which serves as a chaperone binds to pre-secretory proteins and prevents them from folding. SecB then delivers its substrates to SecA, a multi-functional protein that both guides' proteins to the SecYEG channel, and also serves as the ATPase that provides the energy for protein translocation. Following transport through the SecYEG channel, proteins are folded in the periplasm. Periplasmic proteins may be released extracellularly with the help of an additional secretion system (Green and Mecsas, 2016).

3.2.5. Lipoprotein Outer membrane Localization (LOL) Pathway

Many antimicrobial agents that target bacteria are cationic and can interact with the anionic lipid components that are exposed on the bacterial membrane (Epand and Epand, 2011). The LOL system, composed of five proteins, catalyses outer membrane sorting of lipoproteins

(Tokuda, 2009). Lipoproteins present on the surface of Gram-negative bacteria, referred to as surface lipoproteins or SLPs. SLPs are synthesised in the cytoplasm and contain an N-terminal signal peptide. The signal peptide is recognised by the Sec or Tat translocon and the SLP is transported to the periplasm. The SLP molecule is then processed by three enzymes (Lgt/SpII/Lnt) that cleave off the signal peptide and attach a lipid head group. The SLP is then delivered to the OM by the five-component Lol system. Pathogenic *Neisseria* species possess four Lol components where the LolC homodimer compensates for the LolC-LolE heterodimer in the inner membrane (Hooda *et al.*, 2017).

The Lol system comprises an inner-membrane ABC transporter LolCDE complex, a periplasmic carrier protein, LolA, and an outer membrane receptor protein, LolB. Lipoproteins are synthesized as precursors in the cytosol and then translocated across the inner membrane by the Sec translocon to the outer leaflet of the inner membrane, where lipoprotein precursors are processed to mature lipoproteins. The LolCDE complex then mediates the release of outer membrane-specific lipoproteins from the inner membrane, while the inner membrane-specific lipoproteins, possessing Asp at position 2, are not released by LolCDE because it functions as a LolCDE avoidance signal, causing the retention of these lipoproteins in the inner membrane. A water-soluble lipoprotein‑LolA complex is formed as a result of the release reaction mediated by LolCDE (Holland, 2010).

3.3. Protein Secretion Systems in *N. meningitidis*

Gram-negative bacteria have developed several systems for the secretion of proteins across their cell envelope into the extracellular milieu or directly into the cytoplasm of eukaryotic target cells. Six of these systems, designated Types I–VI secretion systems (T1–6SS), are widely disseminated among Gram-negative bacteria. *N. meningitidis* contains only the T1SS and T5SS (Tommassen and Arenas, 2017), as discussed below.

3.3.1. Type 1 secretion system (T1SS)

Type I secretion requires a machinery composed of multimers of three proteins: an OM-embedded channel protein (pore), an IM-embedded ABC (ATP binding cassette) transporter and an adaptor or membrane-fusion protein (MFP) connecting these two components, which is anchored in the IM and has a large periplasmic domain. This machinery spans the entire cell envelope and the secretion involves a single step from the cytoplasm directly into the extracellular medium (Tommassen and Arenas, 2017). Specific substrates are recognized by a C-terminal, non-cleavable motif and pass both membranes in one step (Natale *et al.*, 2008).

T1SS transports one or a few unfolded substrates. These substrates range in function and include digestive enzymes, such as proteases and lipases, as well as adhesins, heme-binding proteins and proteins with repeats-in-toxins (RTX) motifs. T1SS substrates are generally Sec-independent and typically, but don't always, contain a C-terminal signal sequence that is recognized by the T1SS and remains uncleaved. RTX motifs bind to calcium at extracellular, but not intracellular levels. Because calcium binding promotes the folding of these proteins, these large substrates are able to remain unfolded inside the cell (Green and Mecsas, 2016).

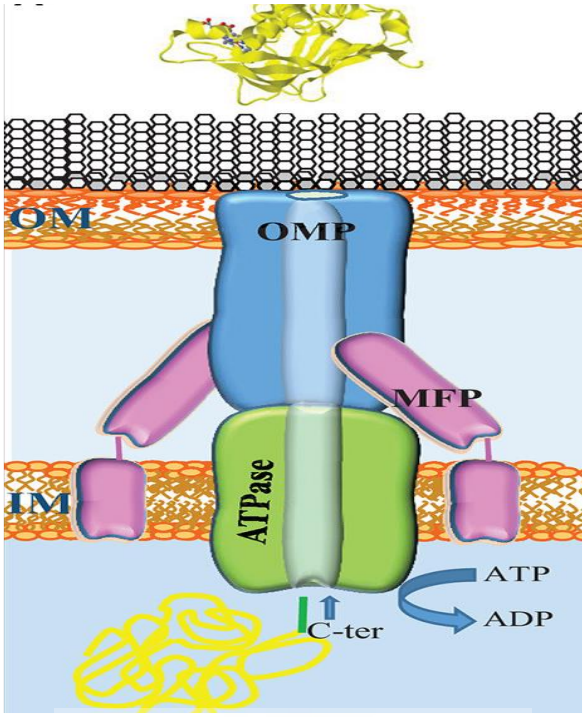


Figure 3.1: Type 1 secretion system (Tommassen and Arenas, 2017)

3.3.2. Type 5 secretion systems (T5SS)

In T5SS, proteins are first transported across the inner membrane via the general export pathway, mediated by the Sec system, after which the periplasmic intermediate is transported across the outer membrane (OM). Based on differences in the latter step, five subsystems are discriminated: T5a–eSS (Grijpstra *et al.*, 2013). Four types encode the translocator (pore-forming) and the passenger (secreted) domains in a single gene: the classical AT (T5aSS), the trimeric AT (T5cSS), the inverted AT (T5eSS), and the fused two-partner system (T5dSS). In two-partner systems (T5bSS), the translocator and passenger are encoded in two separate (typically contiguous) genes (Abby *et al.*, 2016). In *N. meningitidis*, T5aSS, T5bSS, and T5cSS are present (Tommassen and Arenas, 2017).

3.3.2.1. AT (type Va) pathway/ T5aSS: the classical ATs

ATs are classified into monomeric and trimeric ATs. The common principle of all ATs is their dependency on the Sec machinery for IM transit, and the presence of a β-barrel domain

that inserts into the bacterial OM, where it acts as a transporter for the so-called passenger domain(s) destined for surface localization (Leo *et al.*, 2012).

The T5aSS encompasses classical ATs, which are synthesized as precursors consisting of three domains: an N-terminal signal sequence, which is required for targeting the Sec machinery, a passenger domain, and a C-terminal translocator domain. The translocator domain is inserted as a 12-stranded β-barrel into the OM via the general OM-protein assembly machinery, the BAM complex and/or the alternative TAM complex (Tommassen and Arenas, 2017). *Neisseria* can export monomeric AT adhesins, App and MspA/AusI, through a type Va secretion system (Hung and Christodoulides, 2013). The integral translocator domain in the OM has a central hydrophilic channel, which can act as a pore and is essential for transportation of the passenger domain to the cell surface (van Ulsen, 2011). Passenger domains often function as hydrolases, cytotoxins or adhesins and have other activities associated with virulence (Bernstein, 2015).

The *N. meningitidis* NalP protein, the *H. pylori* AlpA protein, the *C. jejuni* CapA protein and the *Bordetella pertussis* SphB1 protein, are exposed on the cell surface by the AT (type Va) pathway (Wilson and Bernstein, 2016).

ATs are produced by a large variety of Gram-negative bacteria. These cell-surface-exposed and secreted proteins have been intensively studied (Grijpstra *et al.*, 2013). In *N. meningitidis*, for example, a homologue of BamB is lacking, but the Bam complex contains another component, RmpM, which stabilizes OMP complexes and is anchored via its periplasmic domain to the peptidoglycan (Grijpstra *et al.*, 2013). Monomeric ATs are expressed as a single polypeptide that contains an N-terminal signal peptide, and the proteins are secreted by the Sec machinery into the periplasm (Leo *et al.*, 2012).

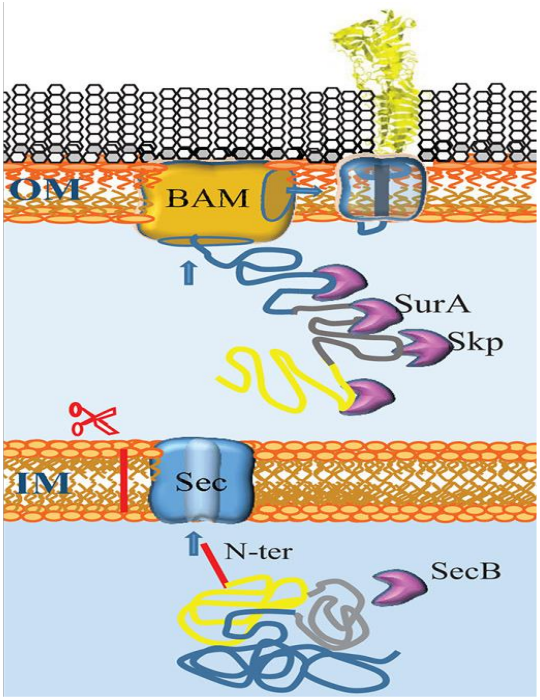


Figure 3.2: AT secretion system (Tommassen and Arenas, 2017)

Following secretion of the unfolded AT protein through the inner membrane, the translocator domain assembles in the outer membrane, forming a 12-stranded β-barrel, usually with the help of a number of accessory factors, including the periplasmic chaperone Skp and the Bam complex (Green and Mecsas, 2016).

3.3.2.2. Two-partner secretion (type Vb) pathway/ T5bSS

In T5bSS, or two-partner secretion (TPS) system, a large β-helical protein, generically designated TpsA, is translocated across the OM via a protein designated TpsB (Jacob-Dubuisson *et al.*, 2013). The two-partner secretion pathway is a branch of type V secretion, alongside the AT. There are two major players in TPS systems, the secreted proteins collectively called TpsA proteins ('the cargos') and their outer membrane partners collectively called TpsB proteins ('the transporters'). TPS systems are dedicated to the secretion across the outer membrane of long proteins that form extended β-helices (Jacob-Dubuisson *et al.*, 2013). TPS system is where a pair of proteins participates in the secretion

process, in which one partner carries the β-barrel domain, while the other partner serves as the secreted protein (Green and Mecsas, 2016).

TPS is a secretion pathway that, in a growing number of Gram-negative bacteria, has been shown to be devoted to the secretion of large virulence-associated proteins (Talà *et al.*, 2008).

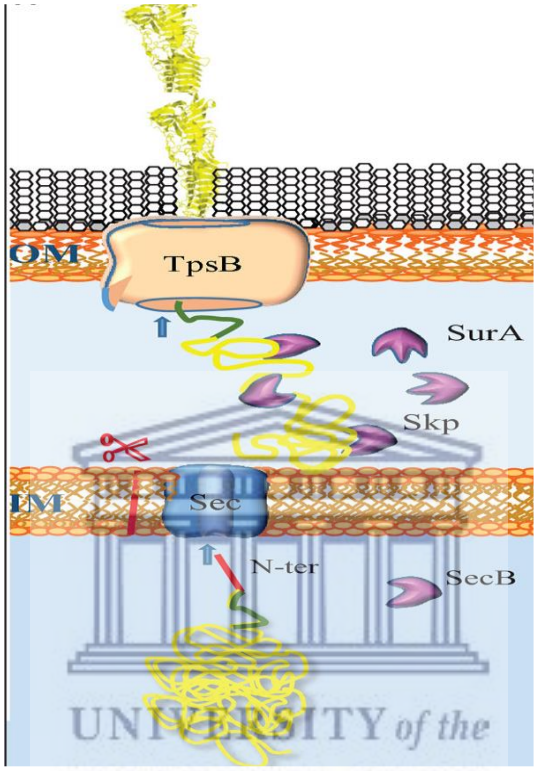


Figure 3.3: The two-partner secretion system (Tommassen and Arenas, 2017)

3.3.2.3. T5cSS: the trimeric ATs

The T5cSS encompasses trimeric ATs, which are similar to classical ATs, but it requires three subunits, which each contribute four β-strands, to form a similar 12-stranded β-barrel as in the T5Ass (Tommassen and Arenas, 2017). In contrast to many of their monomeric counterparts (type Va ATs), they are usually adhesins, do not harbour enzymatic functions and are not released from the cell surface by an autoproteolytic mechanism. Instead, they protrude from the cell surface as relatively rigid rods, with a length of over 250 nm in some cases (Leo *et al.*, 2012). Trimeric ATs follow the same route as type Va ATs for their

biogenesis, the major difference being the presence of three polypeptide chains (Leo *et al.*, 2012).

3.4. Tools for sequence retrieval of *N. meningitidis* receptors

3.4.1. National Center for Biotechnology Information (NCBI)

NCBI (https://www.ncbi.nlm.nih.gov/) is a readily available Web resource, a division of the National Library of Medicine (NLM) at the U.S. National Institutes of Health, is a leader in the field of bioinformatics. NCBI hosts approximately 40 online literature and molecular biology databases including PubMed, PubMed Central, and GenBank that serve millions of users around the world (Tatusova *et al.*, 2013). Over the years the amount and variety of data that NCBI maintains has expanded enormously and can be generally divided into six categories: Literature, Health, Genomes, Genes, Proteins and Chemicals. Each of these six categories has a corresponding web page that lists the relevant databases and tools, along with links to tutorials and other information (Coordinators, 2017).

3.4.2. UniProt Knowledgebase (UniProtKB)

UniProtKB (https://www.uniprot.org/) is a Protein knowledge-base and it consists of two sections: UniProtKB/Swiss-Prot, which is manually annotated and reviewed as well as UniProtKB/TrEMBL, which is automatically annotated and is not reviewed (Boutet *et al.*, 2016). For the retrieval of the receptors, the UniProt/Swiss-Prot section of UniProtKB was used. The UniProt consortium maintains the UniProt KnowledgeBase (UniProtKB), updated every 4 weeks. The Swiss-Prot section of the UniProt KnowledgeBase (UniProtKB/Swiss-Prot) contains publicly available expertly manually annotated protein sequences obtained from a broad spectrum of organisms (Boutet *et al.*, 2016).

The database contains over 60 million sequences, of which over half a million sequences have been curated by experts who critically review experimental and predicted data for each protein. Since the last update in 2014, the number of reference proteomes doubled to 5631, giving a greater coverage of taxonomic diversity (Wu and Consortium, 2016). UniProtKB/Swiss-Prot contains over 550 000 sequences that have been curated by an expert biocuration team. For these entries experimental information has been extracted from the literature and organized and summarized, greatly easing scientists access to protein information (Wu and Consortium, 2016).

3.5. Aims

The aim of this chapter is to search for *N. meningitidis* receptors and linking it to pathways and secretion mechanisms within the bacterium to determine the most significant outer membrane (OM) protein/s for the novel AMP/s (ligand/s) to interact with. Through the interaction of the AMP/s to selected *N. meningitidis* receptors, the bacterium can be detected within biological samples i.e. a diagnostic test for *N. meningitidis*. The objectives are as follows:

- To search for receptors associated to *N. meningitidis* by literature mining.
- To identify the ideal receptors in the OM of *N. meningitidis* to serve as targets for the identified putative AMPs.
- To extract amino acid sequences of the selected receptors using NCBI and UniProt, for 3D modelling and docking studies with the AMPs (chapters 4 and 5).

3.6. Materials and methods

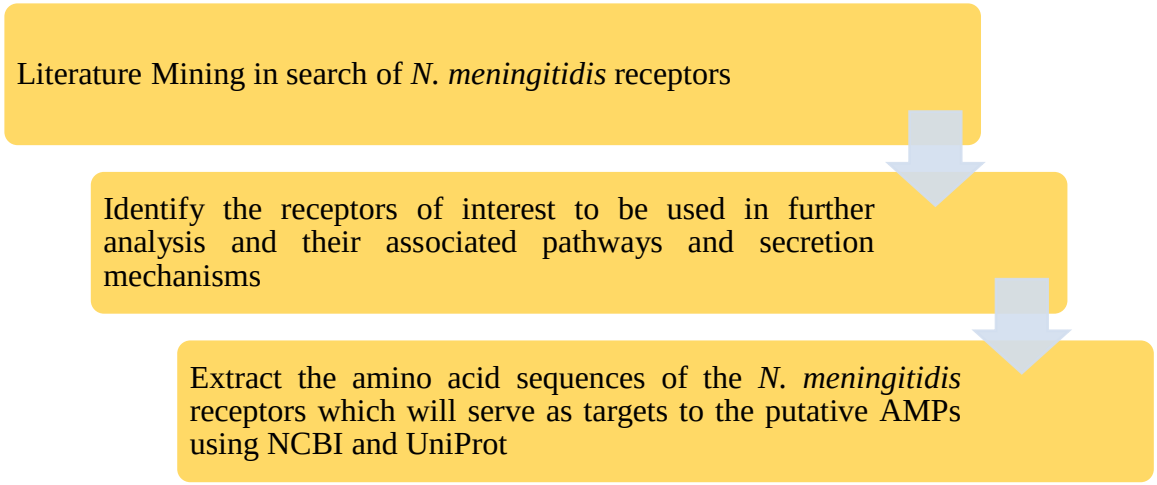


Figure 3.4: Flow chart of the methodology used for identification of receptors as targets to the retrieved AMPs

3.6.1. Biomedical Literature Mining in search of *N. meningitidis* receptors

Various search engines such as Pubmed/NCBI (https://www.ncbi.nlm.nih.gov/pubmed/), Google Scholar (https://scholar.google.co.za/) and Science Direct (https://www.sciencedirect.com/) were accessed with the keywords or key phrases entered into the search engine "cell-surface receptors of *Neisseria meningitidis*", "cell surface proteins of *Neisseria meningitidis*", "outer membrane proteins, receptors of *Neisseria meningitidis*" etc. Literature mining was done using references from published articles, in relation to the above search, a list of Web content results in the form of websites, images, videos or other online data were displayed which was accessed and the relevant information retrieved.

3.6.2. Computational tools for retrieval of amino acid sequences of *N. meningitidis* receptors

The full amino acid sequences of the *N. meningitidis* receptors identified from the previous step were retrieved from National Center for Biotechnology Information (NCBI) database and as confirmation, from the Universe Protein Resource, UniProt Knowledgebase (UniProtKB) database.

NCBI

Using the URL https://www.ncbi.nlm.nih.gov/, the gene name according to the literature articles, e.g. Opc was entered in the query field. The subsequent protein was identified and its ID entered, with *N. meningitidis* as the species / microorganism. The amino acid sequence was extracted and saved for further analysis in the FASTA format

UNIPROT

The extraction of *N. meningitidis* sequences was done by accessing the URL https://www.uniprot.org/. On the home page of UniProtKB, the keyword e.g. "Opc (*N. meningitidis* receptor)" was entered into the "query" box followed by selecting, the Boolean operator "AND" as well as *Neisseria meningitidis*. This was followed by selecting the "Search" button. A list of proteins which matched the keyword was generated. The first result matching the search was selected and under sequence, sequence data in FASTA format was chosen and the sequences copied and saved.

3.7. Results and Discussion

3.7.1. Outer membrane receptors identified for *N. meningitidis* in the literature

Based on literature reviewed, the receptors associated to *N. meningitidis*, have been identified as seen in Table 3.1. The highlighted receptors are discussed more in detail below.

Table 3.1: Summary of *N. meningitidis* Adhesins and Cell Surface Structures identified from the literature

Receptors	Expression
Opc	only expressed by *N. meningitidis*
Opa	expressed by both meningococci and gonococci and some commensal strains
PorA	expressed in all strains of meningococci
PorB	meningococci have both PorA and PorB. *N. lactamica* and *N. polysaccharea*, *N. gonorrhoeae*
NhhA/ Msf	in all meningococcal isolates
NadA	~50% of meningococcal strains, but absent in both *N. gonorrhoeae* and *N. lactamica.*
MspA/Aus1	not present in all meningococcal strains
App	shares ~95% and 73% identity with *N.gonorrhoeae* and *N. lactamica*
TpsA/HrpA - TpsB/HrpB	homology to FHA of *B. pertussis*
GapA-1	in meningococci (>97% identical) and also present in gonococci (99% identical to strain FA1090) and *N. lactamica*
ACP	commensal strains such as *N. lactamica*, *N. polysaccharea* and *N. sicca*
Enolase, DnaK, Peroxidedoxin	moonlighting proteins in various spp.

NHBA/GNA2132	found in several other *Neisseria* species,
RmpM	shares sequence and structural similarity of the OmpA protein from *Escherichia coli*
fHbp	present in most meningococcal strains, a few invasive isolates with either have a frameshifted gene or express fHbp at minimal levels
NaIP/AspA	not all strains of meningococci
TspA	TspA amino acid sequences were also identified in *N. polysaccharea*
Tfp/type IVpilus	expressed diverse set of bacterial species
FBA	meningococcal isolates also 70%, 67% and 65% identical to FBA from *Cupriavidus metallidurans, Xanthobacter falvus* and *Synechocystis spp.*

Meningococcal adhesins can be divided into three broad structural classes, the polymeric hair-like pili, the integral outer membrane proteins (OMP), including the opacity proteins Opa and Opc, which are usually beta barrel structures, and the ATs (including meningococcal serine protease (Msp)A), meningococcal surface fibril (Msf, or Neisseria hia homolog (Nhh)A), and Neisseria adhesin (Nad)A (Hill and Virji, 2012).

Opacity associated proteins being the most abundant adhesins located in the neisserial outer membrane facilitate the interaction of bacteria with a number of host cell types, including epithelial cells on mucosal surfaces and various immune cells, indicating a direct effect on the immune response (Sadarangani *et al.*, 2011).

Opc

The Opc proteins, expressed only by *N. meningitidis* are β-barrelled transmembrane molecules in the outer membrane with the Opa proteins having four, and the Opc proteins five, surface loops (Sa E Cunha *et al.,* 2009, Simonis and Schubert-Unkmeir, 2016). They are basic in nature and target several human receptors of which at least one class of receptors, the heparan sulfate proteoglycans, is recognized by both these proteins particularly on epithelial cells (Sa E Cunha *et al.,* 2009).

Studies on the mechanisms of Opc-mediated interactions with human endothelial cells, integrins were identified as the major receptors at the apical surfaces of the cells. Binding to integrins occurred via a sandwich mechanism in which Opc was shown to first bind to serum-derived integrin receptors, particularly vitronectin and to a lesser extent fibronectin (Figure 3.6), and subsequently form a trimolecular complex with the αvβ3 and α5β1 integrins (Sa E Cunha *et al.*, 2009).

Opa and Opc are similar in size (27–31 kDa) and were initially known as Class 5 proteins (Hill *et al.,* 2010). The crystal structure of Opc was solved in 2002 and has been determined to 2.0 Å resolution showing that this adhesin adopts a ten-stranded *β*-barrel presenting five largely invariant surface-exposed loops (Hill *et al.*, 2010, Hung and Christodoulides, 2013). Opc expressed in *N. meningitidis,* but not *N. gonorrhoeae*, is also an important adhesion (Virji, 2009). (Figure 3.5) Although Opc is not present in every *N. meningitidis* strain, interestingly, epidemiological studies reported meningococcal strains lacking the Opc gene caused severe sepsis with fatal outcomes, but not meningitis, suggesting that Opc could play a key role in the induction of meningitis in IMD (Simonis and Schubert-Unkmeir, 2016).

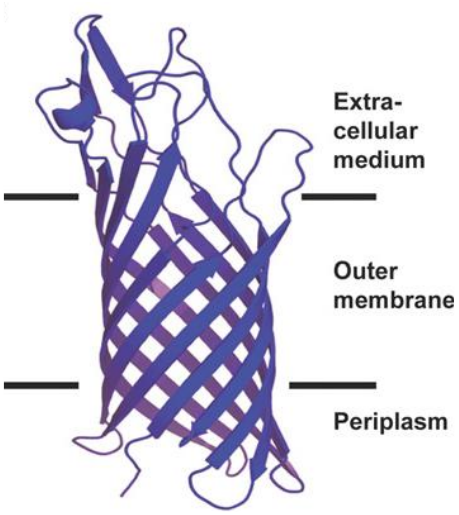


Figure 3.5: Structure of *N. meningitidis* Opc protein. This Figure was extracted from (Hill *et al.*, 2010).

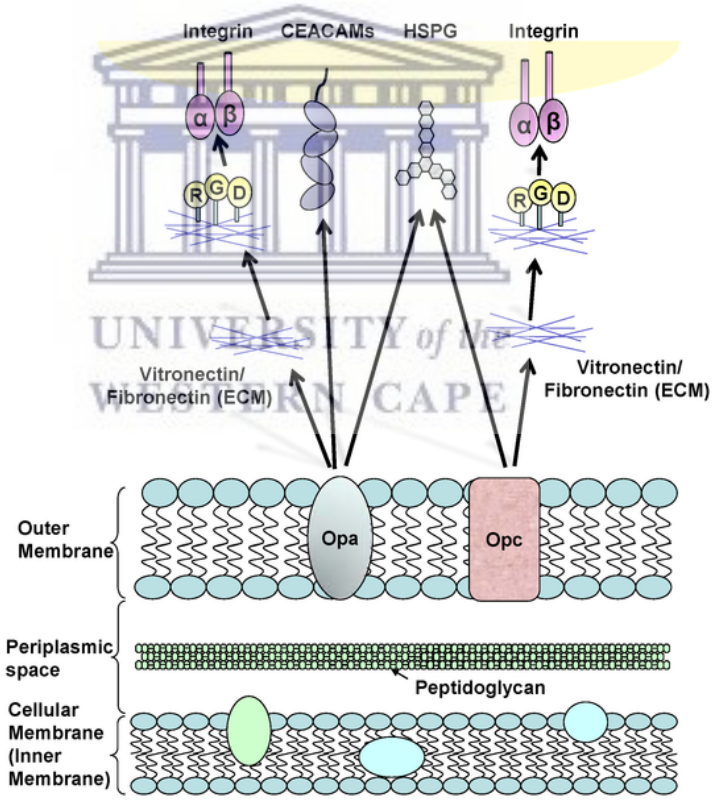


Figure 3.6: Interactions between *Neisseria* Opa and Opc proteins and human host cells involves multiple binding receptors (Hung and Christodoulides, 2013).

Porin (Por)

N. meningitidis expresses two distinct porins, PorA (formerly class 1 protein) and PorB (formerly class 2/3 protein based on molecular mass). Both porins are β-barrel proteins, which associate into trimers in the bacterial outer membrane through which small hydrophilic nutrients diffuse into the cell. Individual porins vary in molecular mass with PorA (~46 kDa) being expressed in all strains of meningococci (Hill *et al.*, 2010). Porins comprise up to 60% of the proteins present in the *Neisseria* OM. PorA and PorB of *N. meningitidis* are trimeric voltage-gated pores that mediate ion exchange between the organism and its environment (Hung and Christodoulides, 2013; Peak *et al.*, 2016).

AT Adhesin: NhhA

NhhA, N*eisseria H*ia/Hsf *h*omologue, is an outer membrane protein homologous to the Hia and Hsf adhesins of *Haemophilus influenza.* These proteins exhibit divergent functional properties and often contribute to bacterial adherence, invasion, micro colony formation, and transepithelial trafficking or serum resistance. The meningococcal NhhA protein is an AT adhesin that is present in all tested meningococcal strains (Sjölinder *et al.*, 2008).

In a study by Scarselli *et al.*, 2006, it was demonstrated that NhhA facilitates bacterial attachment to host cells *in vitro*. Sjölinder *et al.*, 2008 demonstrated that the protein is essential for bacterial colonization of the nasopharyngeal mucosa and NhhA has a determining role in protecting bacteria from host innate immune defences, which subsequently affects the outcome of the disease process. Taken together, the results of this study revealed a multifaceted impact of NhhA during the development of meningococcal disease.

A Bioinformatic analysis of meningococcal outer-membrane adhesins by Andreae *et al.*, 2018 revealed that Msf (Msf; also referred to as Neisseria hia homologue A, NhhA) is found

in all meningococcal isolates, and displays diversity in the N-terminal domain. The Meningococcal surface fibril (Msf), binds to the activated form of vitronectin (Vn) (Figure 3.7) to increase *N. meningitidis* survival in human serum (Hill *et al.*, 2015). The C-terminal domain of NhhA is able to form a translocator domain that allows localization of the protein to the bacterial surface (Sjölinder *et al.*, 2012).

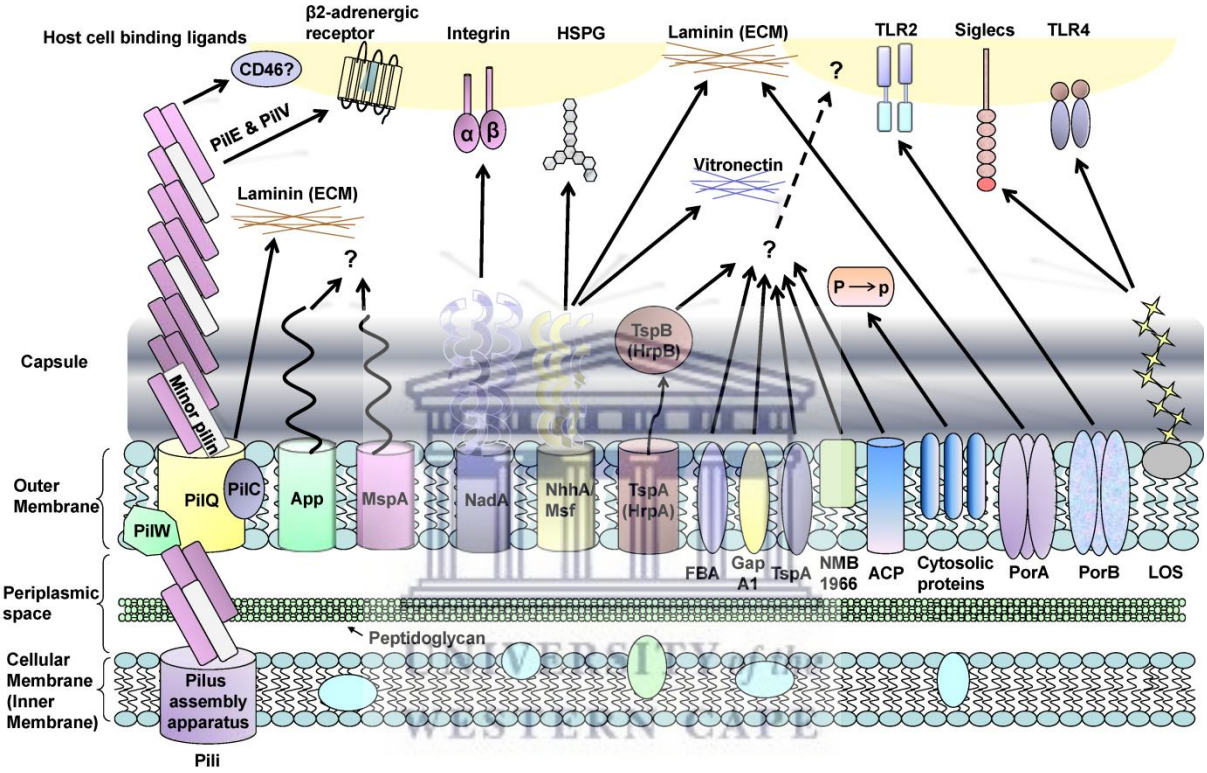


Figure 3.7: Schematic review of *N. meningitidis* surface molecules and their interactions with human host cell binding ligands (Hung and Christodoulides, 2013)

3.7.2. *N. meningitidis* receptors chosen to interact with the AMPs

The selection of receptors was based on specificity to *N. meningitidis* and meningococcal strains thus allowing for a potential LFD to identify *N. meningitidis* specifically as well as other meningococcal strains. Binding studies within this research will determine the AMP/s to be used within a potential LFD.

Table 3.2: Selected *N. meningitidis* receptors to use as targets to the putative AMPs

Receptors	Sensitivity
Opc	expressed only by N. meningitidis
NhhA	OMP present in all meningococcal strains
PorA	expressed in all strains of meningococci

3.7.3. Retrieval of receptor sequences

3.7.3.1. NCBI sequences of *N. meningitidis* receptors selected

The retrieved sequence of the receptor NhhA, resulted in 591 amino acids (aa), Opc 272 aa and PorA 392 aa. The sequences of the *N. meningitidis* proteins extracted from NCBI are shown in Appendix B: Table B.1, B.3 and Table B.5.

3.7.3.2. UniProt identification of *N. meningitidis* amino acid sequences of chosen receptors

As in the NCBI database the search of UniProt resulted in identical sequences of NhhA 591 aa, although it was extracted from the *N. meningitidis* serogroup B (strain MC58), it is the only sequence reviewed and records with information extracted from literature and curator-evaluated computational analysis as seen in Figure 3.8.

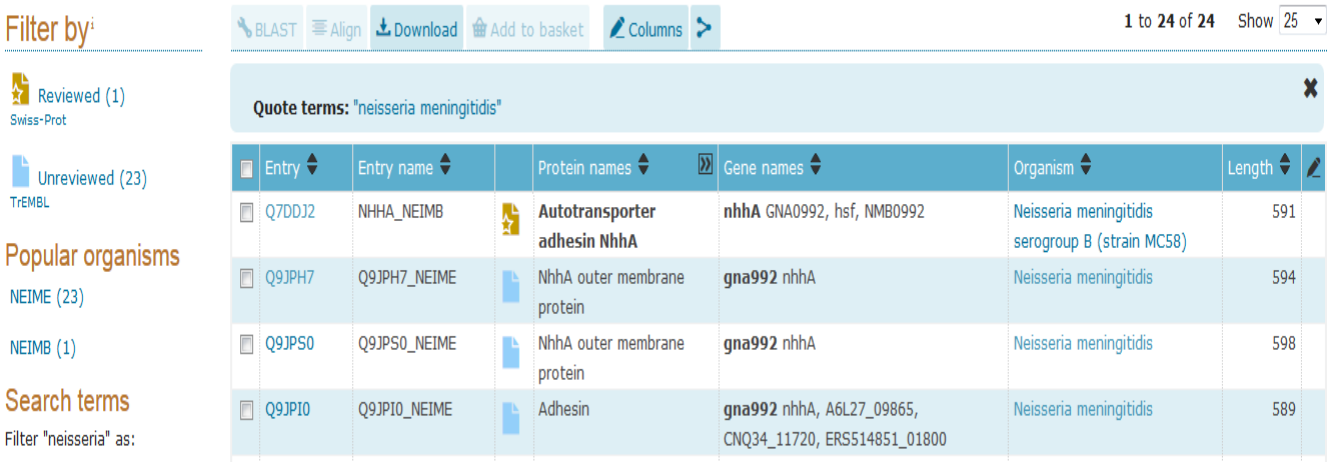


Figure 3.8: Taken from UniProt displaying the Reviewed result for NhhA

The Opc 272 aa, sequence extracted were unreviewed for the organism *N. meningitidis*. The PorA 392 aa sequences were reviewed for the *N. meningitidis* serogroup B (strain MC58) as well. The sequences of the *N. meningitidis* proteins extracted from UniProt are shown in Appendix B: Table B.2, B.4 and Table B.6.

3.8. Summary

This chapter focused on the identification of *N. meningitidis* receptors and linking these receptors to pathways and secretion mechanisms to determine the most significant OM protein for the novel AMPs to interact with through literature mining. The choice of receptor to be targeted by the AMP/s are based on its unique expression within *N. meningitidis* such that a diagnostic test using AMP/s selectively identify this bacterium within a patient sample.

Particularly in a case were an individual has been infected by various strains of meningococci it will be of great importance that the more virulent strains are identified for the best treatment options.

Two databases were used, for cross referencing of the extracting amino acid sequences of the three selected receptors, namely NCBI and UniProt which yielded identical amino acid sequences for the selected proteins.

Reasons and justifications why certain OMPs where not selected over the three *N. meningitidis* proteins (NhhA, Opc and PorA) for further use in this study; as the aim of this study is to detect *N. meningitidis* with high specificity only.

1) The conservation of NadA, fHbp, and NHBA might suggest that they are not naturally immunologically exposed (during colonising infection) and as such might not be as immunogenic as more variable outer membrane proteins (Sadarangani and Pollard, 2010). The fHbp gene is present in strains of commensal *Neisseria* species that are closely related to *N. meningitidis* (Seib *et al.,* 2015).
2) Neisserial App (Adhesion and penetration protein) shares a high degree of homology to Hap (Haemophilus adhesion and penetration protein, the product of the *hap* gene) in *Haemophilus influenza.* All *Neisseria* species possess the app gene and the meningococcal App protein amino acid sequence shares ~95% and 73% identity with *N. gonorrhoeae* and *N. lactamica* App, respectively (Hung and Christodoulides, 2013).
3) Glyceraldehyde 3-phosphate dehydrogenases (GAPDHs) are cytoplasmic glycolytic enzymes, with *N. meningitidis* having two genes, gapA-1 and gapA-2, encoding GAPDH enzymes (Tunio *et al.*, 2010). GapA-1 is surface-located and highly conserved in meningococci (>97% identical) and also present in gonococci (99% identical to strain FA1090) and *N. lactamica* (90% identical to strain ST640) (Tommassen and Arenas, 2017).
4) Hung *et al.*, 2013 demonstrated adhesin complex protein (ACP), as a protein located on the surface of meningococci and expressed by patient and carriage strains, including *N. sicca, N. polysaccharea,* and *N. lactamica*
5) NHBA is a surface-exposed lipoprotein that is a target of both meningococcal and human proteases (Sadarangani and Pollard, 2010). NHBA gene is ubiquitous in

meningococcal strains of all different serogroups and it has also been found in several other *Neisseria* species, including *N. lactamica, N. polysaccharea and N. Flavescens* (Serruto *et al.,* 2010).

6) The C-terminal domain of rmpM belongs to a family that shares sequence and structural similarity with the C-terminal domain of the OmpA protein from *Escherichia coli* (Maharjan, *et al*., 2016).
7) *N. meningitidis* is the only species of the genus *Neisseria* that possesses a functional *nalP* gene/protein, although *nalP* is not present in any commensal species and is present in the human pathogen *Ngon* (Oldfield *et al.,* 2013).
8) Similar TspA amino acid sequences were also identified in *N. polysaccharea*, but not in *N. lactamica* or *Ngon* (Hung and Christodoulides, 2013).
9) Shams *et al.*, 2016 demonstrated that FBA is present on the surface of pathogenic and nonpathogenic species of *Neisseriae*.
10) Opacity-associated (Opa) protein is commonly expressed in both meningococci and gonococci. Expression of Opa proteins has been demonstrated in some commensal strains, including *N. lactamica, N. subflava* and *N. flavescens* (Hung and Christodoulides, 2013).

In conclusion, only proteins PorA, NhhA and Opc will be selected for further studies to identify the ideal receptor to bind the AMPs for accurate and specific detection of MD.

Chapter 4

Physicochemical Characterization of the putative AMPs and predicting the 3D structures of the Anti-*N. meningitidis* putative AMPs and *N. meningitidis* receptors

4.1. Introduction

Although AMPs are a diverse group of molecules in terms of sequence, structure and sources, there are several properties that are common to almost all AMPs. AMPs display a net positive charge ranging from +2 to +13, are ≈10–40 amino acids long and may contain a specific cationic domain. The cationic nature can be attributed to the presence of basic amino acid residues lysine and arginine (and sometimes histidine) with the presence of hydrophobic residues alanine, leucine, phenylalanine or tryptophan (greater than 30%) being common, and other residues such as isoleucine, tyrosine and valine also present (Schmidtchen *et al.,* 2014; Travkova *et al.,* 2017; Lee *et al.,* 2017; Kumar *et al.,* 2018). Hydrophobic interactions are among the most important driving forces in nature and in biology. These interactions determine the structure of proteins and cells as well as the self-assembly of membranes. The hydrophobic effect plays a key role in ligand binding processes (Schauperl *et al.,* 2016).

Amphipathicity which is shared by all antimicrobial peptides refers to the relative abundance of hydrophilic and hydrophobic residues or domains within the AMPs. It can be thought of as the balance between the cationic and hydrophobic residues, not just at the primary sequence level, but also in terms of the 2D or 3D structures of the AMPs (Kumar *et al.,* 2018). Amphipathicity is very important for binding to microbial membranes and can be achieved via a multitude of peptide conformations (Travkova *et al.,* 2017).

The initial event in attachment is the adsorption onto the outer membrane surface of the microbe, which can occur within tens of nanoseconds and is largely mediated by electrostatic interactions between cationic AMPs and anionic lipopolysaccharides (LPS) molecules. Characteristics of the interaction of AMPs with the outer and cytoplasmic membranes include: length of the AMP sequence, the total and density of cationic charges, the total number of hydrogen bond donors and the 3D conformation of the AMP in solution and at the membrane (Kang *et al.,* 2012; Mihajlovic and Lazaridis, 2012; Schmidtchen *et al.*, 2014; Juba *et al.,* 2015; Li *et al.,* 2017). Hence the electrostatic and hydrophobic interactions are two driving forces that steer an AMP toward and into the bacterial membrane (Li *et al.,* 2017).

In a study by Mihajlovic and Lazaridis 2010 they used molecular dynamic simulations to investigate binding preferences of antimicrobial peptides and showed that four AMPs (alamethicin, melittin, a magainin analogue, MG-H2, and piscidin 1) bind strongly to membrane pores.

The physicochemical properties are directly derived from the peptide sequence and comprise a complete set of parameters that accurately describe AMPs. Torrent *et al.,* 2011 concluded that sequence derived parameters are enough to characterize antimicrobial peptides. Various tools for physicochemical characterisation have been created and are embedded in databases such as APD and Bactibase.

AMPs are found to exist in a wide range of secondary structures such as α-helices, β-strands with one or more disulphide bridges, loop and extended structures (Pushpanathan *et al.,* 2013; Schmidtchen *et al.,* 2014; Travkova *et al.,* 2017). *In silico* modelling reveals that both the alpha helical and beta-sheet conformations are amphiphilic (Li *et al.,* 2017). Torrent *et al.,* 2011 observed that AMPs tend to be randomly coiled in solution, with a low tendency to

present any defined structure. Many AMPs form α-helices, particularly when interacting with lipid membranes (Schmidtchen *et al.*, 2014).

The shape and the physicochemical properties on the protein molecular surfaces govern the specific molecular interactions in protein-ligand complexes. Therefore, studies as diverse as those on protein folding, protein conformational stability, inter- and intra- protein interactions, molecular recognition and docking; as well as applications-orientation, such as drug design, protein and peptide solubility, crystal packing and enzyme catalysis, benefit from an accurate and precise representation of the molecular surfaces (Nicolau Jr *et al.*, 2014).

Great progress has been made in structure determination of proteins/peptides using experimental methods, such as X-ray crystallography, high-resolution electron microscopy and nuclear magnetic resonance (NMR) spectroscopy, these approaches are generally still expensive, time consuming, and not always applicable. Computational methods for predicting the 3D structures of proteins enjoy a high degree of interest and are the focus of many research and service development efforts (Schwede *et al.*, 2008). I-TASSER is one such a program for 3D structure determination and will be used in this study.

4.2. Physicochemical Parameters of the putative anti-*N. meningitidis* AMPs

4.2.1. Antimicrobial Peptide Database (APD)

The APD is a comprehensive database for peptide discovery, nomenclature, classification, information search, calculations, prediction and design of AMPs. This database can be accessed at the website (http://aps.unmc.edu/AP). To include the calculation and prediction section of APD the following URL must be accessed at (http://aps.unmc.edu/AP/prediction/prediction_main.php). The host defense AMPs registered in APD cover the five kingdoms (bacteria, protists, fungi, plants, and animals) or three

domains of life (bacteria, archaea, and eukaryota) (Wang *et al.,* 2016). In the peptide property calculations section, in addition to peptide length, net charge, amino acid composition and Boman index previously found in APD, and in an updated version, APD3 enables the calculation of molecular weight, molecular formula, molar extinction and coefficient. APD also provides identification of the most similar sequences. Once a new peptide is sequenced, one would like to know which known sequences it most resembles, which can be conducted in the prediction interface of APD (Wang *et al.,* 2016).

The Prediction interface allows users to input a new peptide sequence. The program will carry out a residue analysis on the peptide. It also predicts whether the new peptide has the potential to be antimicrobial based on some known principles. In terms of structure, only some simple predictions can be made. For instance, when the hydrophobic residues appear every two to three residues in the peptide sequence, an amphipathic helix will be predicted (Wang and Wang, 2004). However, I-TASSER will be used for structure prediction in this study (see 4.3.1.).

4.2.2. Bactibase

The Bactibase database is a data repository of bacteriocin natural antimicrobial peptides freely available at (http://bactibase.pfba-lab.org); a web-based platform enabling easy retrieval, via various filters, of sets of bacteriocins that will enable detailed analysis of a number of microbiological and physicochemical data for both Gram-positive and Gram-negative bacteria (Hammami *et al.,* 2007).

4.3. *De novo* modelling of the 3D Structure of *N. meningitidis* receptors and Putative Anti-*N. meningitidis* AMPs

De novo (or *ab initio*) methods aim to predict the structure of a protein purely from its primary sequence, using principles of physics that govern protein folding and/or using

information derived from known structures but without relying on any evolutionary relationship to known folds (Schwede *et al.,* 2008). The term *ab initio* prediction often refers to the subset of *de novo* methods that rely on energy functions based solely on physicochemical interactions. Such approaches, using full-atom simulations with empirical force fields as well as explicit and implicit solvent models, have been successful in predicting the folding of short peptides (Schwede *et al.,* 2008). The popularity of the *de novo* (or *ab initio*) method means that "threading" became a generic term to describe carrying out protein fold recognition (McGuffin, 2008).

Fold recognition and threading techniques are a comparatively fast and inexpensive way to build a close approximation of a structure from a sequence, without the time and costs of experimental procedures (McGuffin, 2008). The structural biology of proteins is much more complex, where each protein has its own unique 3D structure. Since small changes in the sequence of a protein can have strong effects on its biophysical properties, experimental determination of protein structures is a laborious and often unpredictable endeavour. The computational modelling of a protein's structure has therefore attracted substantial interest in the field of bioinformatics to complement experimental structural biology efforts to characterize the protein universe (Schwede, 2013). *De novo* modelling methods do not explicitly rely on whole known structures as templates. Thus, the structure of any protein can be predicted by these *de novo* methods (Schwede *et al.,* 2008).

4.3.1. I-TASSER (*I*terative *T*hreading *ASSE*mbly *R*efinement)

The iterative threading assembly refinement (I-TASSER) server is an integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm. Starting from an amino acid sequence, I-TASSER generates three-dimensional (3D) atomic models from multiple threading alignments and iterative structural

assembly simulations (Roy *et al.*, 2010). The I-TASSER server is freely available to the academic community at http://zhang.bioinformatics.ku.edu/I-TASSER (Zhang, 2008).

The community-wide Critical Assessment of Structure Prediction (CASP) experiments have been designed to obtain an objective assessment of the state-of-the-art in the field, where I-TASSER was ranked as the best method in the server section of the recent 7th CASP experiment and has been ranked as the best method for the automated protein structure prediction in the last two CASP experiments (Zhang, 2008)

For each submitted sequence, the following items are returned by email after I-TASSER modeling to estimate the accuracy of the I-TASSER predictions: (1) up to five predicted models ranked based on the structure density of the SPICKER clustering (algorithm to identify the near-native models from a pool of protein structure decoys); (2) Confidence score (C-score) of all the I-TASSER models; (3) estimated Template Modelling score (TM-score) and Root Mean Square Deviation (RMSD) for the first model in the form of Estimation ± Deviation (Zhang, 2008).

The C-score is an estimate of the quality of the predicted models and is normally in the range [−5, 2] and a model of C-score >−1.5 usually has a correct fold, with TM-score >0.5. The biological functions of the protein, including ligand-binding sites, enzyme commission number, and gene ontology terms, are then inferred from known protein function databases based on sequence and structure profile comparisons (Yang and Zhang, 2015).

A large-scale benchmark test demonstrates a strong correlation between the C-score and the TM-score of the first models with a correlation coefficient of 0.91 (Zhang, 2008, Yang *et al.*, 2015). Combining C-score and protein length, the accuracy of the I-TASSER models can be predicted with an average error of 0.08 for TM-score and 2 Å for RMSD (Zhang, 2008).

Identifying template proteins from solved structure databases which have a similar structure or similar structural motif as the query sequence, with TM-score <0.17 meaning random predictions and TM-score >0.5 meaning correct topology for all sizes of proteins. As a consequence of the sensitivity of TM-score on structural topology, it was found that the correlation coefficient of C-score and TM-score (0.91) is much higher than that of C-score and RMSD (0.75). The estimation of TM-score is usually more reliable than that of RMSD for the I-TASSER models, i.e. TM-score estimation has usually a much smaller systematic error than RMSD estimation (Roy *et al.,* 2010).

As the biological function of protein molecules is determined by their 3D shape (which dictates how the protein interacts with receptors or other protein molecules), one of the most common motivations for predicting the protein structure is to use the structural information to gain insight into the protein's biological function (Roy *et al*., 2010).

The users can view the structures of these files with any professional molecular visualization software e.g., PyMOL and RasMol and draw customized figures for various purposes (Yang and Zhang, 2015).

4.4. Visualization of the 3D Structure of *N. meningitidis* receptors and Putative Anti-*N. meningitidis* AMPs

4.4.1. PyMol

PyMol is a free cross-platform molecular graphics system made possible through recent advances in hardware, internet, and software development technology. PyMOL has been released under a completely unrestrictive open-source software license so that all scientists and software developers can freely adopt PyMOL and then distribute derivative works based on it without cost or limitation. PyMOL supports most of the common representations for macromolecular structures: wire bonds, cylinders, spheres, ball-and-stick, dot surfaces, solid

surfaces, wire mesh surfaces, backbone ribbons, and cartoon ribbons (DeLano, 2002). Originally designed to: (1) visualize multiple conformations of a single structure [trajectories or docked ligand ensembles] (2) interface with external programs, (3) provide professional strength graphics under both Windows and Unix, (4) prepare publication quality images, and (5) fit into a tight budget (DeLano and Bromberg, 2004).

PyMOL can be downloaded for free via the internet at http://www.pymol.org and currently runs on a variety of platforms: Windows, Linux, IRIX, Mac OSX, and Tru64 Unix (DeLano, 2002). Visualization is essential to understanding structural biology. Only open−source software allows you to surmount problems by directly changing and enhancing the way software operates, and it places virtually no restrictions on your power and opportunity to innovate (DeLano and Bromberg, 2004).

4.5. Aims

The aim of this chapter was to determine whether the peptide sequences identified by HMMER conform to known AMPs both in physicochemical characteristics as well as structure. The objectives were to:

- Characterise the predicted peptide sequences based on their physicochemical properties to ensure that they conform to known AMPs using APD and Bactibase
- Predict the 3D structures of the *N. meningitidis* receptors and the putative AMPs using I-TASSER server
- Visualise the 3D structures generated using the PyMOL 1.3. Software

4.6. Materials and methods

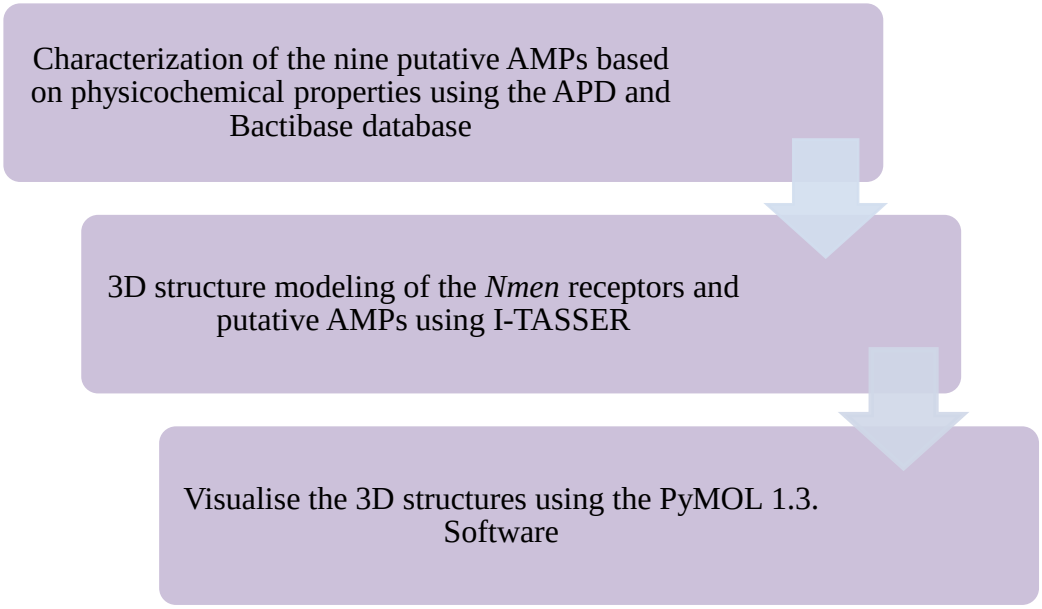


Figure 4.1: Flow chart of the methodology used in Chapter 4

4.6.1. Determination of the Physicochemical Parameters of the Putative Anti-*N. meningitidis* AMPs

The physicochemical properties studied of the nine identified putative anti-*N. meningitidis* AMPs included: (i) the number of residues, (ii) total hydrophobic ratio, (iii) total net charge, (iv) the Isoelectric point, (v) the Boman Index (or protein binding potential), (vi) molecular weight, (vii) the instability index of the proteins, (viii) Arginine (Arg) % and Lysine (Lys) %, (ix) other amino acids with high %, using the calculation and prediction section in APD and in Bactibase and is presented in Table 4.1.

4.6.1.1. Physicochemical Properties using APD

The APD database was used in this chapter for the calculation and characterization of the nine putative anti-*N. meningitidis* AMPs. All nine putative AMP sequences were submitted to the AMP characterization software Antimicrobial Peptide Database (calculations & prediction tab) (http://aps.unmc.edu/AP/prediction/prediction_main.php) to determine the

characteristics of each putative AMP. Using the prediction interface of APD, the sequence of each AMP separately were copied and placed under the "Please input your peptide sequence *(one-letter code for the standard 20 amino acids and no space"* instruction then 'Submit' was selected.

4.6.1.2. Physicochemical Parameters Analysis within Bactibase

All nine putative AMP sequences were submitted to the AMP characterization software Bactibase (physiochemical properties tab) (http://bactibase.pfba-lab- tun.org/ physicochem) to determine the characteristics of each putative AMP. Using the Tools tab, physicochemical profile was selected and each physicochemical profile of the query sequence was analysed by submitting the nine AMPs sequences individually. Extraction of some physicochemical properties from Bactibase, which were not found in APD were all combined in Table 4.1.

4.6.2. Predicted 3D *de novo* structures of the Anti-*N. meningitidis* Putative AMPs and *N. meningitidis* receptors by I-TASSER

The amino acid sequence of the *N. meningitidis* receptors and nine AMPs to be modelled were inputted in the FASTA format and submitted to the online server of the I-TASSER software (http://zhang.bioinformatics.ku.edu/I-TASSER) and the result generated in PDB format.

Depending on the protein size, the I-TASSER modelling procedure takes a maximum of 48 hours (typically 5–10 hours for a sequence around 200 residues). After the modelling is finished, an email is sent, which include the PDB format files of up to 5 predicted models, C-score of the models, and the predicted RMSD and TM score of the first model. A brief explanation of the RMSD, TM-score, and C-score is also provided in the email (Zhang, 2008).

The 3D structure prediction of the *N. meningitidis* receptors (NhhA, Opc and PorA) was done using the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) online database. The amino acid sequences of the receptors were copied into the input box, 'run I-TASSER' was selected to submit the job. All output files were saved for visualization of the structures for the *N. meningitidis* receptors.

4.6.3. Visualization using PyMol software.

The PyMol software was accessed and in the 'File' menu, a PDB file was selected corresponding to the AMP or receptors 3D structure as generated by I-TASSER. Hide lines and show cartoons were selected. The structure was enhanced and saved in two formats the PSE and PNG formats.

4.7. Results and Discussion

4.7.1. Characterization of the putative AMPs based on Physicochemical Properties in APD

The physicochemical properties of the nine putative anti-*N. meningitidis* AMPs were determined to ensure that these peptides conform to other known AMPs. The output of the APD database results were (refer to Table 4.1):

a) Number of amino acids (residues)

AMPs are characterized by a short length; they generally comprise less than 50 amino acids. This property minimizes the probability of being degraded by bacterial proteases (Osorio *et al.*, 2015). All AMPs were short in length ranging from 27-34 amino acids thus falling within the range expected for AMPs.

b) Total hydrophobic ratio

The hydrophobicity is an important stabilization force in protein folding. It is considered to be the driving force of the peptide to the core of the bacterial membrane (Osorio *et al.*, 2015). It was ascertained that all the putative anti-*N. meningitidis* peptides have hydrophobic values that are above 30%, which is the anticipated value for hydrophobicity content of an AMP (Table 4.1). The presence of hydrophobic residues (greater than 30%) in AMPs is common (Lee *et al.*, 2017). In a study by Chen *et al.*, 2007, they investigated the role of hydrophobicity in the antimicrobial activity of a synthetic V13KL AMP by systematically decreasing or increasing the hydrophobicity and noted that decreasing AMP hydrophobicity was associated with reduced antimicrobial activity. The hydrophobic and positively charged domains of indolicidin, a small antimicrobial peptide with 13 amino acids isolated from bovine neutrophils, rich in tryptophan (39%) and arginine (23%) residues, are crucial for its interactions with bacterial pathogens (Mojsoska and Jenssen, 2015). Although extremely variable in length, amino acid composition and secondary structure, all peptides can adopt a distinct membrane-bound amphipathic conformation (Nguyen *et al.*, 2011).

c) Total net charge

The characterisation of the putative anti-*N. meningitidis* AMPs based on their net charge showed AMPs with a positive total net charge ranging from +1 to +8, contributed to by the high percentage of the presence of the positively charged amino acids Lysine and Arginine (see Table 4.1) within the respective AMP sequences. Although YYNN4 had 0% Lys and YYNN9 had 0% Arg, it had high Arg and Lys percentages respectively. It was shown that out of the nine anti-*N. meningitidis* AMPs, all AMPs have a high net positive charge except for YYNN4, YYNN3 and YYNN9, which showed a low positive net charge and this, can be attributed to the absence of the positively charged amino acid Lysine within the AMP sequence YYNN4, Arg in YYNN9 and low % Lys in YYNN3. AMPs have a positive net charge at pH 7, which provides binding specificity to the negatively charged bacterial

membranes through electrostatic interactions (Osorio *et al.,* 2015). The antimicrobial activity of most of the members of the defensin family appears to be related to their cationicity. Human defensin 5 as an example, interacts with the bacterial surface via its arginine residues and thus exerts its antimicrobial activity. Replacement of arginine residues at position 9 and 28 with alanine or lysine residues reduces the antibacterial killing as well as the host cell interaction (Mojsoska and Jenssen, 2015). The antimicrobial peptide, human neutrophil peptide-1 (net charge +3), appeared to be more effective against *S. aureus* than human β-defensin-3 (net charge +11) thus cationicity alone cannot account for selectivity. There is a large insertion of Arg and Lys residues at the C terminus of human β-defensin-3 and significant bactericidal activity against Gram-positive *Staphylococcus aureus* at physiological salt concentrations (0.154μM) (Schibli *et al.,* 2002; Mojsoska and Jenssen, 2015). The lethal concentration of human neutrophil peptide-1 for killing *S. aureus* NCTC 8530 is 0.8 ± 0.2μM (Varkey and Nagaraj, 2005).

Many linear AMPs are unstructured in an aqueous solution and require a membranous environment to adopt a stable amphipathic conformation. Membrane interactions remain important even for intracellular- targeting peptides because they must have a means of translocation. The cationicity of the AMPs also promotes interactions with negatively charged moieties on other biomolecules such as outer membrane lipids, nucleic acids and phosphorylated proteins (Nguyen *et al.,* 2011).

d) Molecular weight (MW)

The molecular weight is directly related to the length of the amino acid sequence and is expressed in units called daltons (Da). AMPs due to its short length are characterized by a molecular weight <10 kDa (10000 Da) (Osorio *et al.,* 2015). As seen in Table 4.1, all AMPs

molecular weight represented in Da, shows results with low MW in line with the expected range.

e) Boman Index

The Boman index estimates the potential for one protein to interact or bind to different receptors of a pathogen. AMPs tend to not interact with other proteins (the proposed mechanism of action is based on the interaction with membranes), so their Boman indices are usually less than zero or near 0 (Osorio *et al.,* 2015). An antimicrobial peptide with a Boman index value lower or equal to 1 kcal/mol signifies that the peptide will likely exhibit high antimicrobial activity with no side effects and a Boman index value of 2.50 to 3.00 will indicate multifunctional activity with hormone-like activities (Boman, 2003).

In other words, a high Boman index value indicates that an AMP will play a variety of different roles within the cell due to its ability to interact with a wide range of proteins (Azad *et al.*, 2011). The potential protein interaction is an easy way to differentiate the action of antimicrobial peptides (protein-membrane) through this index (Osorio *et al.*, 2015).

The putative Anti-*N. meningitidis* AMPs, YYNN2, YYNN3, YYNN4 YYNN8 and YYNN9 has a potential to bind to other proteins based on their Boman indices that are less than 2.5 kcal/mol. It was observed for YYNN 2, YYNN8 and YYNN 9, a Boman index less than zero, meaning that they may be good antimicrobial peptides. The Boman indices of the remaining putative AMPs were high, thus indicating that the peptide could be multifunctional with hormone-like activities.

Three peptides, namely LL-37, PR-39 and VIP with index values 2.48 or higher are predicted to have a high binding potential. VIP is regarded as a neurotransmitter and the multifunctional data for the other two would justify these peptides as potential hormones. The cecropin-melittin hybrid CA(1–7)M(2–9) was designed only to be antibacterial and it has a

negative index (−0.54). Of the natural peptides, magainin has the lowest index (0.42). It predicts a low potential for interaction with receptors with a relatively good correlation between the index and the known properties of the respective peptides (Boman, 2003).

4.7.2. Characterization of the putative AMPs based on Physicochemical Properties in Bactibase

As seen in Table 4.1 the following additional physicochemical parameters were provided by Bactibase:

a) Isoelectric point

The isoelectric point (*pI)* is the pH value at which the net charge of a molecule is zero (Kozlowski, 2016). It is a variable that affects the solubility of the peptides under certain conditions of pH. When the pH of the solvent is equal to the pI of the protein, it tends to precipitate and loose its biological function (Torrent *et al.,* 2011). AMPs and non-AMPs have similar average isoelectric points of 9.26 and 9.20, respectively. A high positive net charge is required for AMPs, whereas it does not represent a distinctive feature in non-AMPs, probably due to the diverse functions exerted by these peptides (Torrent *et al.,* 2011). The isoelectric point is significant since it represents the pH where solubility is typically minimal.

Overall, the net charge of the peptide is strongly related to the pH, as seen in table 4.1, where AMPs YYNN2, YYNN7 and YYNN8 shows a *pI* of above 11 and a high positive net charge. Although, seen in YYNN9 a low net charge is displayed with a *pI* of 11.1, this can be due to diverse functions exerted by antimicrobial peptides.

b) Instability Index

The instability index, an estimate of peptide stability, is based on its amino acid composition (Osorio *et al.,* 2015). Antimicrobial peptides tend to be considered stable with index values

less than 40 and peptides with values greater than 40 are said to be unstable (Wang and Wang, 2004). The instability index of AMPs could predict the peptide to be stable or unstable in an *in vivo* environment, since in protein-peptide interaction biomolecules are in three-dimensional conformation, thus requiring a stable form. The findings seen in table 4.1, only YYNN 2, YYNN8 and YYNN 9 had an instability value less than 40 and the rest of the peptides instability values were over 40.

4.7.3. Predicted *in-silico* 3D structures of the putative AMPs and *N. meningitidis* receptors

The amino acid sequences of the putative anti-*N. meningitidis* AMPs and *N. meningitidis* receptors were submitted to I-TASSER for 3D structure determination and visualized using PyMol.

Based on Table 4.2, the C-score, TM-score and RMSD of the putative anti-*N. meningitidis* peptides and the three *N. meningitidis* receptors, 3D structures obtained were that of a good model. As seen in Figure 4.3, the I-TASSER results indicated that all the predicted structures of the putative anti-*N. meningitidis* peptides contained majorly α-helical structures.

a) C-Score

A model generated with a C-score >−1.5 usually has a correct fold, although an estimate of the quality of the predicted models and is normally in the range [−5, 2] (Yang and Zhang, 2015). It was shown that the predicted 3D structures of NhhA, Opc and PorA had a C-score of -0.73, -0.17 and -0.72, respectively. Although all the putative anti-*N. meningitidis* AMPs gave C-score values higher than -1.5, except YYNN2 having a C-score of -1.76. YYNN2 C-score is still within range, which indicated a predicted model of good quality.

Table 4.1: Physicochemical properties for the nine putative anti-*N. meningitidis* AMPs as determined by APD (Wang *et al.*, 2016) and Bactibase (Hammami *et al.*, 2007).

Antimicrobial Peptide Database (APD)									Bactibase			
AMP	Residues	Total hydrophobic ratio	Total net charge	Molecular weight	Protein-binding Potential (Boman index)	Arginine %	Lysine %	Other amino acids with high %	Most Common amino acid	Isoelectric point	Instability Index	Job ID number - I-TASSER
YYNN1	33	39%	5	3828.493	3.36 kcal/mol	21%	3%	Ser 15%; Cysine 12%	R	10.00	78.33	S396810
YYNN2	33	51%	8	3724.599	0.82 kcal/mol	9%	15%	Ala, Val, Ile 12%	K	11.79	19.27	S397251
YYNN3	33	42%	2	3872.584	2.28 kcal/mol	15%	3%	Ser, Cys, Ile, Leu 9%	R	8.51	53.2	S397304
YYNN4	34	41%	1	3803.33	2.42 kcal/mol	14%	0%	Ser 11%	R	8.04	42	S397545
YYNN5	33	39%	4	3836.472	3.15 kcal/mol	18%	3%	Ser, Cys 12% Leu 9%	R	9.11	88.21	S397779
YYNN6	33	39%	4	3800.436	3.08 kcal/mol	18%	3%	Ser 15% Cys 12% Leu 9%	R	9.11	81.88	S397819
YYNN7	33	42%	6	3784.483	3.04 kcal/mol	21%	3%	Ser 15% Cys 12% Leu 9%	R	11.17	72.49	S397983
YYNN8	27	51%	6	3067.751	0.65 kcal/mol	7%	14%	Ile, Val 11% Ala 14%	KA	11.84	38.61	S398034
YYNN9	27	48%	3	2529.055	-1.07 kcal/mol	0%	11%	Gly 25% Leu 22% Ala 11%	G	11.1	-15.53	S398202

b) TM-Score

A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity (Roy *et al*., 2010).

All *N. meningitidis* proteins and putative anti-*N. meningitidis* AMPs had TM-scores greater than 0.5, meaning that the proteins have correct topology or structural shape (Roy *et al*., 2010), except YYNN2 which had a TM-score of +/- 0.50, which meant that the template modelling structure used for its structure prediction was not similar to the peptide which structure was to be predicted (Roy *et al*., 2010).

c) RMSD

High-resolution models with (RMSD) values in the range of 1–2 Å usually meet the highest structural requirements and are sometimes suitable for computational ligand-binding studies and virtual compound screening. Medium-resolution models, roughly in the RMSD range of 2–5 Å, can be used for identifying the spatial locations of functionally important residues such as active sites and the sites of disease-associated mutations. Finally, even models with the lowest resolution, from an otherwise meaningful prediction, i.e. models with an approximately correct topology, predicted using either *ab initio* approaches or based on weak hits from threading, have a number of uses including protein domain boundary identification (Roy *et al.,* 2010).

The RMSD of YYNN3, YYNN4 of the predicted structures had a value between 1-2 Å and the RMSD of YYNN8 and YYNN9 of the predicted structures had a value between 2-5 Å. The RMSD of YYNN1, YYNN5, YYNN6 and YYNN7 had a value less than 1 Å and the three *N. meningitidis* receptors and YYNN2 of the predicted structures had a value greater than 5 Å (see Table 4.2).

Table 4.2: Quality evaluation scores of the predicted 3D structures by I-TASSER

AMP Name and *N. meningitidis* receptors	C-score	Exp.TM-Score	Exp.RMSD (Å)
YYNN1	0.98	0.85+-0.08	0.5+-0.5
YYNN2	-1.76	0.50+-0.15	5.3+-3.4
YYNN3	-0.03	0.71+-0.12	1.9+-1.6
YYNN4	0.56	0.79+-0.09	1.0+-1.0
YYNN5	1	0.85+-0.08	0.5+-0.5
YYNN6	0.99	0.85+-0.08	0.5+-0.5
YYNN7	0.98	0.85+-0.08	0.5+-0.5
YYNN8	-0.69	0.63+-0.14	2.8+-2.0
YYNN9	-1.15	0.57+-0.15	3.6+-2.5
NhhA	-0.73	0.62+-0.14	9.4+-4.6
Opc	-0.17	0.69+-0.12	6.3+-3.9
PorA	-0.72	0.62+-0.14	8.4+-4.5

For the three *N. meningitidis* receptors, a high RMSD value is observed, but with a good or reasonable TM-score. This often happens when the protein is big. In these situations, the user should judge the quality of the predicted model based on the expected TM-score rather than the expected RMSD (Roy *et al.*, 2010).

As a consequence of the sensitivity of TM-score on structural topology, Roy, Kucukural *et al.*, 2010 found in a benchmark test that the correlation coefficient of *C*-score and TM-score (0.91) is much higher than that of C-score and RMSD (0.75). Therefore, the estimation of TM-score is usually more reliable than that of RMSD for the I-TASSER models, i.e., TM-score has usually a much smaller systematic error than RMSD in the estimation.

Following physicochemical characterization, Yang and Zhang, 2015 have rigorously evaluated the algorithms in the I-TASSER server in community-wide blind experiments and demonstrated considerable advantages of this method over peer methods in protein structure and function prediction. With numerous feedback from the user community, a variety of new developments have been made to the server to improve the quality of the server in atomic-

level structure refinement, structure-based function annotation, local quality estimation and user interface communication (Yang and Zhang, 2015). Based on these constant and current method developments, this database provided an accurate and excellent structure and functions prediction method for the novel AMPs in this study.

4.7.4. Visualisation of 3D structures output

The AMPs showed secondary structures including α-helices and extended shapes. YYNN1, YYNN3 YYNN4, YYNN5, YYNN6 and YYNN7 all have similar structures, represented by an extended partial α-helical structure or loop structure with partial α-helical secondary structure. The AMPs YYNN2, YYNN8 and YYNN9 exhibited secondary structures, represented by α-helical structure. It is noted that none of the putative AMPs resulted in β-sheet structures. Many AMPs form α-helices, particularly when interacting with lipid membranes (Schmidtchen *et al.,* 2014). AR-23 is a melittin-related peptide with 23 residues that has a high α-helical amphipathic structure which results in strong bactericidal activity and cytotoxicity (Zhang *et al.,* 2016).

Opc has a β-barrel structure with five surface-exposed loops shown by the results obtained by I-TASSER server. According to the literature, the structure of Opc is a β-barrel protein also with five surface-exposed loops (Hill *et al.,* 2010).

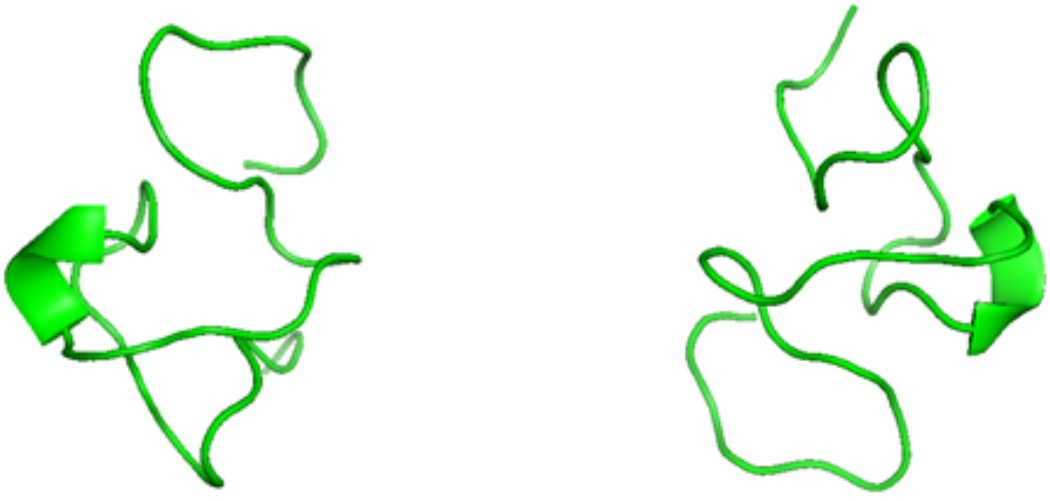
PorA (formerly class 1 protein) are β-barrel proteins, which associate into trimers in the bacterial outer membrane through which small hydrophilic nutrients diffuse into the cell. Individual porins vary in molecular mass with PorA (~46 kDa) being expressed in all strains (Hill *et al.,* 2010). The same structure was predicted by I-TASSER for this protein (seen in Figure 4.2).

According to the literature, NhhA is a trimeric autotransporter adhesin (TAA) family of secreted Gram-negative OM proteins and is capable of trimerization to form a complete β-barrel (Hung and Christodoulides, 2013).

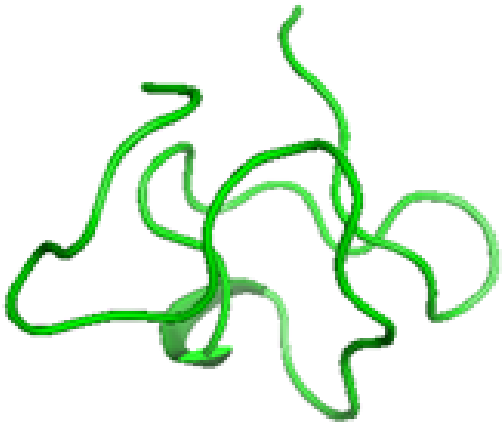
The NhhA structure was predicted with good topology although showing a low C-score value, indicating a lack of good templates in the protein structure library, this is seen in the structural difference of the prediction by the I-TASSER server, where NhhA was not predicted as a trimeric β-barrel.

Ab initio modeling of medium-to-large size proteins without using templates is a major challenge in the field (Roy *et al.,* 2010).

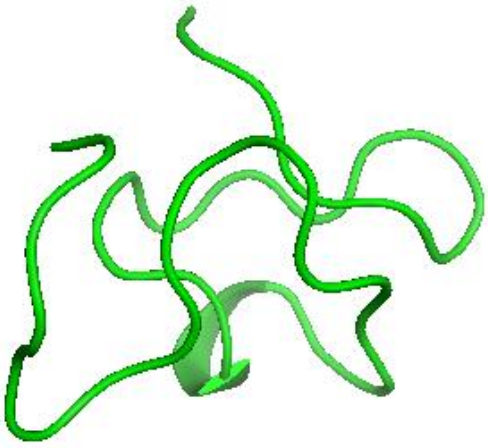
YYNN1 **YYNN2**



YYNN3



YYNN4

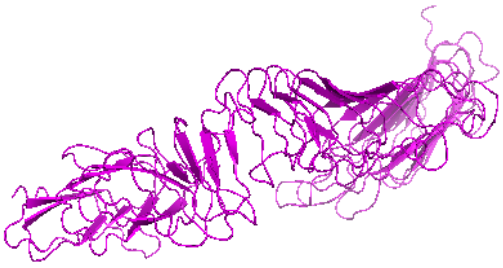


YYNN5

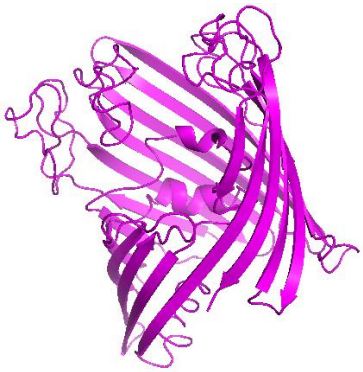
YYNN6

YYNN7

YYNN8



YYNN9 **NhhA**



Opc **PorA**

Figure 4.2: 3D structures of putative anti-*N. meningitidis* AMPs (YYNN1, YYNN2, YYNN3, YYNN4, YYNN5, YYNN6 YYNN7, YYNN8 and YYNN9) and *N. meningitidis* receptors (NhhA, Opc and PorA) as predicted by I-TASSER.

4.8. Summary

The aim of this chapter was to determine that the peptide sequences identified by HMMER (Chapter 2) conform to known AMPs both in physicochemical characteristics as well as 3D structure. Their physicochemical properties were predicted using APD and Bactibase. The predication of the 3D structures of the putative AMPs as well as that of the *N. meningitidis* receptors were carried out using the I-TASSER server and visualized using the PyMOL 1.3. Software.

The results obtained from APD and Bactibase indicates that the identified peptides conform to known AMPs based on their physiochemical characteristics. In addition, the 3D structures of the putative anti-*N. meningitidis* AMPs and the receptors of *N. meningitidis* were of good quality; based on their C-score, TM-score and RMSD value as determined by I-TASSER. The putative AMPs that stood out with good structural prediction based on C-score, TM-score and RMSD and considered most suitable for computational ligand-binding studies were

YYNN3 and YYNN4. I-TASSER results indicated that all the predicted structures of the putative anti-*N. meningitidis* peptides contained majorly α-helical structures with the *N. meningitidis* protein structures obtained, the same as seen in the literature.

Chapter 5

In Silico Site-Directed Mutagenesis Study

5.1. Introduction

Site-directed mutagenesis (SDM) is a powerful tool for discovering the importance of an amino acid in the function of the protein. Changes in amino acid type can reveal sites that are important in maintaining the structure of the protein. Conversely, when investigating functionally interesting sites, it is important to choose replacement residues that are unlikely to affect structure dramatically (Betts and Russell, 2007). Experimental or computational SDM can be applied to optimize the specificity of a peptide for its target (Vanhee *et al.,* 2011).

Experimentally, SDM like alanine scanning mutagenesis and experimental biochemical identification methods, such as crystallographic complex determination, are costly and time-consuming. Instead, computational methods predict 'hotspots' residues with higher efficiency and lower cost (Tuncbag *et al.,* 2009; Aumentado-Armstrong *et al.,* 2015; Qiao *et al.,* 2018; Wang *et al.,* 2018). 'Hotspot' residues of proteins are fundamental interface residues that help proteins perform their functions. In the binding interface, 'hotspots' are packed significantly more tightly than other residues and are key in understanding binding mechanisms and the stability of protein-protein interactions. The formation of biological complexes, such as protein-based complexes, is generally accomplished by the presence of single residues with high binding affinity (Tuncbag *et al.,* 2009; Ramos and Moreira, 2013; Cukuroglu *et al.,* 2014).

The binding affinity and the specificity of the protein-protein interactions are provided by energetically important residues. Mutations of these residues cause dissociation of the proteins or force them to change their binding modes (Cukuroglu *et al.,* 2014). *In silico*

mutagenesis of interaction interfaces has revealed that peptide interfaces contain 'hotspot' residues, reminiscent of those found in protein–protein interfaces (Vanhee *et al.*, 2011).

In silico SDM is simple, fast, has a low computational cost and can be applied to a wide range of proteins providing a correct anatomic image of an interface. It can be used prior to an experimental investigation helping in the 'hotspot' detection and the choice of the amino acids to mutate (Moreira *et al.*, 2007), potentially increasing the binding affinity between the receptor and the peptide/ligand.

Probably the most common broad division of amino acids is those that prefer to be in an aqueous environment (hydrophilic) and those that do not (hydrophobic). The latter can be divided according to whether they have *aliphatic* or *aromatic* side chains (Betts and Russell, 2007). In SDM, consideration of the physical and chemical properties of the amino acids guides the choice of replacements, along with knowledge of the structure of the protein (Betts and Russell, 2007).

Consequences of substitutions were examined as shown by Betts and Russell, 2007. When considering a mutation, it is important to consider how conserved the position is within other homologous proteins. Conservation across all homologues (paralogues and orthologues) should be considered carefully. These amino acids are likely to play key structural roles or a role in a common functional theme (i.e. catalytic mechanism). Other amino acids may play key roles only in the particular orthologous group (i.e. they may confer specificity to a substrate), thus meaning they vary when considering all homologues (Betts and Russell, 2007). Darnell *et al.*, 2007 proposed prediction of 'hotspots' with computational alanine scanning method (using atomic contacts, physicochemical properties and shape specificity contributions), as a hybrid computational model combining decision tree. The technique of SDM has been used to characterize gene and protein structure–function relationships,

protein–protein interactions, binding domains of proteins, or active sites of enzymes for decades (Carrigan *et al.,* 2011).

Computational approaches for the prediction of interaction sites are based on an attempt to identify general features that are shared by many interaction sites and then use these features to identify new putative interaction sites (Ofran, 2009). The machine-learning approaches try to learn the complicated relationship between 'hotspot' and various residue features and then distinguish hot spots from the interface residues (Wang *et al.,* 2018).

There has been considerable interest applying machine-learning methods to predict 'hotspots' such as Neural Networks (Ofran and Rost, 2007), Decision Trees (Darnell *et al.,* 2007; Darnell *et al.,* 2008), Support Vector Machines (Zhu and Mitchell, 2011) and Random Forests (Wang *et al.,* 2012), for developing computational prediction methods to complement the mutagenesis experiments.

Several algorithms have been developed for prediction and include energy-based methods, such as Robetta and FOLDEF, which make a prediction based on an estimate of the energetic contribution to binding for every interface residue. Also, knowledge-based methods that try to learn the complex relationship between 'hotspots' and various residue features in training data and then predict new 'hotspots'(Wang *et al.,* 2012).

Overall, computational approaches have become a valuable complement to experimental approaches and can reduce the number of mutations that experimental researchers have to pursue when attempting to establish principles about binding mechanisms (Xia *et al.,* 2016). For the purpose of this work the next section will focus on KFC for the prediction of 'hot spot' residues.

5.2. Knowledge-based FADE and Contacts (KFC) Server

The Knowledge-based FADE and Contacts (KFC) Server is a web-based machine learning approach for the prediction of 'hotspots' and is accessible at http://kfc.mitchell-lab.org (Darnell *et al.*, 2008). K-FADE uses shape specificity features calculated by the Fast Atomic Density Evaluation (FADE) program, and K-CON uses biochemical contact features. The combined KFADE/CON (KFC) model displays better overall predictive accuracy than computational alanine scanning (Robetta–Ala) (Darnell *et al.*, 2007; Chen *et al.*, 2013), analysing the features of experimentally determined 'hotspot' residues from protein complexes having known structures. Knowledge-based models are created from this data to predict 'hotspot' residues in new protein complexes (Darnell *et al.*, 2008).

A machine learning algorithm analyses these features and produces a model for predicting new 'hotspots'. K-FADE and K-CON predict several 'hotspot' that are missed by Robetta–Ala, further illustrating that knowledge-based methods improve the scope and accuracy of 'hotspot' predictions within protein interfaces. KFC method is considerably faster than Robetta–Ala. By joining forces, KFC and Robetta–Ala (KFCA) together predict nearly three-quarters of experimentally observed 'hotspot' residues. The server is organized into three main sections: the submission page, the queue and the job viewer. On the submission page, the user provides a protein structure and defines the interface to be analysed. Next, the submitted job enters the server's queue for processing. Afterwards, the job viewer superimposes the results from the KFC analysis onto the protein structure. These tools help the user to quickly analyse a protein interface and to simply visualize the structural environment around putative hot spots (Darnell *et al.*, 2008).

5.3. Aim

The aim of this chapter was to use the parental AMPs as templates to generate derivative AMPs that display increased predicted binding affinity for the NhhA, Opc and PorA proteins using site directed mutagenesis.

- Identify 'hotspot' residues responsible for the interaction between the receptors and the parental AMPs using the KFC server
- Perform site directed mutagenesis (SDM) by changing ‘non hotspot’ amino acid residues within the binding interaction area to increase the binding affinity of the AMPs to their interacting proteins
- Determine the physicochemical characteristics of each derivative AMP using Bactibase and APD
- Predict the 3D structures of the derivative AMPs using the algorithm I-TASSER

5.4. Materials and Methods

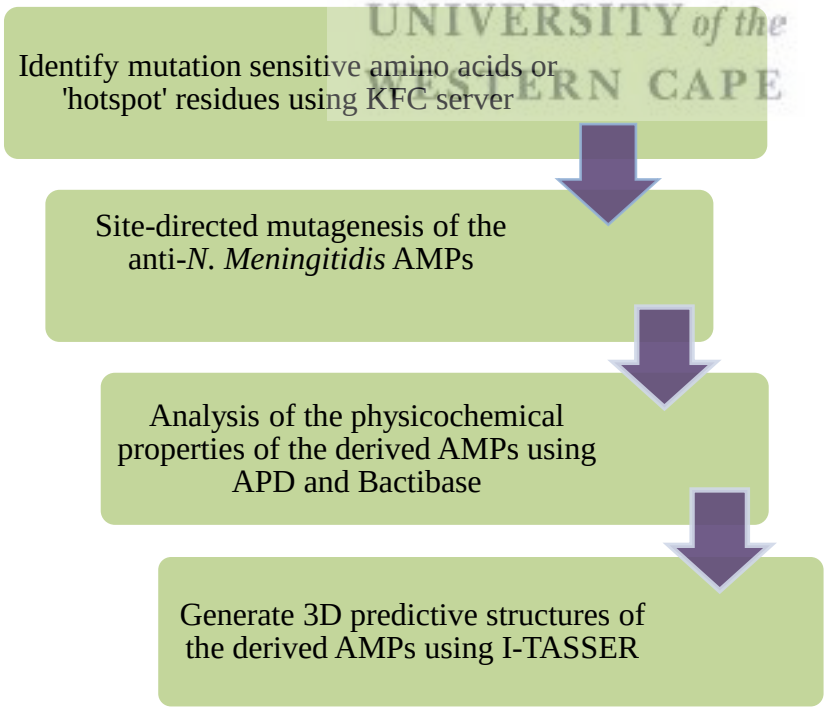


Figure 5.1: Outline of Methodology for mutated AMPs

5.4.1. Identify 'hotspot' residues or mutation sensitive residues within the parental AMPs

Before *in silico* site-directed mutagenesis, essential amino acids have to be identified**.** 'Hotspot' residues were predicted using the KFC server on each of the parental AMPs in relation to their binding to each identified receptor (NhhA, Opc and PorA).

5.4.2. Selection of positions and residues for mutagenesis

The substitutions of amino acids (non hotspots) introduced into the parental AMP sequences were done to increase the binding affinity when bound to *N. meningitidis* receptors. All amino acid substituted to generate mutated AMPs where of similar characteristics to the amino acids present in the parental AMPs, as to maintain the predicted structure and functioning of the AMPs.

5.4.3. Determination of the Physicochemical Parameters of the mutated AMPs

Following site directed mutagenesis (SDM) using KFC of the parental (wild-type) identified AMPs (see Chapter 2) it was important to ascertain that the mutated AMPs retains the same functionality as the parental AMPs. Analysis of the mutated AMPs was carried out as described in Chapter 4 section 4.5.1.1. using Antimicrobial Peptide Database (APD) (Wang and Wang, 2004; Wang *et al.,* 2009; Wang, 2015; Wang *et al.,* 2016) and Bactibase (Hammami *et al.,* 2007; Hammami *et al.,* 2010). The physicochemical properties of the derived anti-*N. meningitidis* AMPs can be seen in Table 5.5.

5.4.4. 3D structure prediction of the mutated AMPs

The amino acid sequences of the derivative AMPs were used as input for I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER). Predicted structure outputs were given as

PDB files, with a scoring which represented the accuracy of each structure prediction. The visualisations of the 3D structures were done using the PyMOL 1.3. Software.

5.5. Results and Discussion

5.5.1. 'Hotspot' identification

For comparisons of side chain functional properties, each of the 20 common amino acids were grouped by their standard side chain class as (1) acidic, (2) aliphatic, (3) amidic, (4) aromatic, (5) basic, (6) hydroxylic, or (7) sulfur-containing. For comparison of molecular dimensions, standard molar mass in grams per mol (g/mol) were assigned to each of the 20 common amino acids and was used as a surrogate for molecular dimensions as amino acids with larger side chains have greater molar mass relative to amino acids with smaller side chains. Requiring one or more key amino acids to share either the same side chain classification or molar mass as the template species was used for determination of chemical susceptibility in order to produce conservative predictions, as dramatic differences among amino acid residues are more likely to change the protein-chemical interaction relative to minor differences (Doering *et al.*, 2018).

The outputs of the task from KFC server resulted in amino acid residues that contribute to the interaction of the three *N. meningitidis* proteins (NhhA, Opc and PorA) and the AMPs, with the "hotspot" residues within the binding area (seen in Table 5.1 as an example). Results represent a confidence score of K-FADE prediction (worst value is 0 and its best value is 1) and a confidence score of K-CON prediction (worst value is 0 and its best value is 1) (Table 5.1). Figure 5.2 shows the interaction between AMP YYNN1 and the *N. meningitidis* Opc protein, indicating amino acid residues "hot spots" within this interaction.

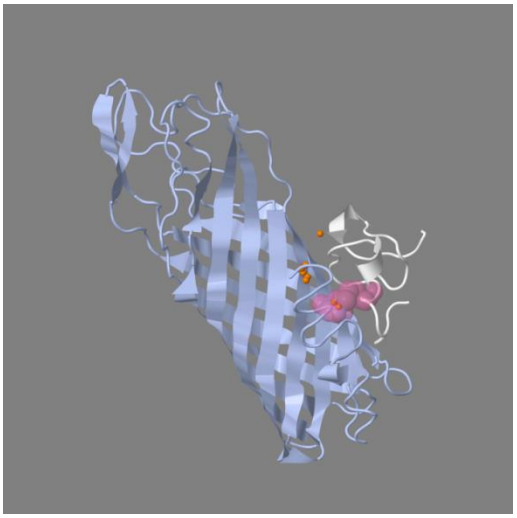


Figure 5.2: Displays the Interaction of AMP YYNN1 and the *N. meningitidis* Opc protein. The small orange balls highlight local regions of well-matched shape specificity between the binding partners. The translucent surface around the residues (pink) displays the area where a predicted 'hotspot' is situated in this interaction (E.g. ARG8 for the AMP and THR7 for the receptor).

'Hotspot' residues were not mutated, since this is where the interaction and strong binding occurs. These residues contribute most of the energy involved in protein-protein or protein-peptide interactions. By mutagenesis of the "hot spot" residues, the binding affinity will be decreased instead of increased.

A result of hotspots of the AMP YYNN1 to the *N. meningitidis* protein, NhhA can be seen in the Table 5.1.

Table 5.1: 'Hotspot' Prediction RESULT from KFC server for AMP YYNN1 interacting with NhhA protein

NhhA-YYNN1					
Amino acid Name	Residue number	Predicted K-FADE class	K-FADE Conf	Predicted K-CON class	K-CON Conf
TRP	2	Hotspot	0.35	Hotspot	0.08
ARG	3	-------	-2.5	-------	-0.81
GLY	4	-------	-0.91	-------	-0.7
VAL	5	-------	-1.44	-------	-0.88
SER	6	-------	-1.14	-------	-0.88
LEU	7	-------	-0.1	-------	-0.08
ARG	8	Hotspot	1.01	Hotspot	0.34
PRO	9	Hotspot	0.72	-------	-0.12
ILE	10	-------	-2.18	-------	-0.66
GLY	11	-------	-1	-------	-0.66
ALA	12	Hotspot	0.89	-------	-0.47
SER	13	-------	-1.96	-------	-0.52
ARG	15	Hotspot	0.31	Hotspot	0.14
ASP	16	Hotspot	1.22	-------	-0.13
ASP	17	-------	-0.06	-------	-0.43
SER	18	Hotspot	1.23	-------	-0.28
GLU	19	Hotspot	1.34	Hotspot	0.19
ARG	23	-------	-1.25	-------	-0.53
CYS	25	-------	-2.27	-------	-0.69
ARG	28	-------	-0.32	-------	-0.17
ARG	29	-------	-1.91	-------	-0.81
CYS	30	-------	-0.04	-------	-0.24
SER	31	-------	-0.72	-------	-0.7
LEU	32	-------	-2.19	-------	-0.96

KFC predicts computational 'hotspots' by a machine learning approach by structural features such as atomic contacts and H-bonds and gives a binary answer whether a residue is a 'hot spot' or not (Darnell *et al.,* 2008). For prediction on large-scale data, KFC is preferred because of its computational effectiveness and comparable performance (Tuncbag *et al.,* 2009).

Unravelling 'hotspots' in binding interfaces continues to stimulate interest, since reliable prediction of key residues in the interface has immediate applications in protein engineering and it is an attractive alternative therapy for many diseases (Moreira *et al.,* 2007).

5.5.2. Site directed mutagenesis

After the identification of the 'hotspots', position of amino acid residues were studied and mutated by changing 'non hotspot' residues. Most amino acid residue substitution is based on

increasing the binding affinity of the ligand within the complex with its receptor. The criteria of mutation of a specific amino acid was based on similarity of the substituted amino acid with that of the parental molecule, since the amino acids has similar physicochemical properties, it was expected that the mutated AMPs will have increased binding affinity, following site-directed mutagenesis.

YYNN1a (S6T) mutation was made by substituting the serine residue of the parental peptide with the threonine amino acid, since both are hydrophilic and neutral, hydroxyl-containing side chains with similar molecular weight. This concept was used throughout the mutation process of amino acid substitution of parental AMPs. See Table 5.2, 5.3 and 5.4 for all the substitutions made within the parental AMPs. The following nomenclature will be used subsequently in this thesis: a - the respective AMPs bound to NhhA, b - the respective AMPs bound to Opc and c - the respective AMPs bound to PorA e.g. AMPs bound to NhhA will be designated YYNN1a – YYNN9a.

Table 5.2: Displaying the position of each amino acid on the parental AMPs and the amino acid substitution on that same AMP, which would still bind selectively to NhhA protein.

NhhA-YYNN	
Mutated AMPs	**Mutation**
YYNN1a	S6T
YYNN2a	L26V
YYNN3a	P9V
YYNN4a	E19L
YYNN5a	G4V
YYNN6a	Q15C
YYNN7a	T22M
YYNN8a	N19C
YYNN9a	T26R

Table 5.3: Displaying the position of each amino acid on the parental AMPs and the amino acid substitution on that same AMP, which would still bind selectively to Opc protein

Opc-YYNN	
Mutated AMPs	**Mutation**
YYNN1b	E19Y
YYNN2b	F5W
YYNN3b	S18R
YYNN4b	Y27F
YYNN5b	H15K
YYNN6b	E19Y
YYNN7b	A12L
YYNN8b	N12Q
YYNN9b	S5T

Table 5.4: Displaying the position of each amino acid on the parental AMPs and the amino acid substitution on that same AMP, which would still bind selectively to PorA protein

PorA-YYNN	
Mutated AMPs	**Mutation**
YYNN1c	C14M
YYNN2c	L26I
YYNN3c	G4V
YYNN4c	S13K
YYNN5c	G11L
YYNN6c	D17E
YYNN7c	G4A
YYNN8c	R2H
YYNN9c	V25L

5.5.3. Physicochemical properties of the derived anti-*N. meningitidis* AMPs

The physiochemical properties of the derived AMPs were determined using APD and Bactibase to ascertain that the mutated AMPs still conform to known AMPs

The physicochemical properties remained fairly similar following substitution based site-directed mutagenesis (results shown in Table 5.5). It could be observed from Table 5.5 that

the parameters used for physiochemical characterization of the derived AMPs are the same as those used for the parental AMPs. Slight changes in physicochemical characteristics post site-directed mutagenesis were observed for the mutated AMPs. This observation can be explained by the amino acid substituted within the derived AMPs. Although similar in nature to the amino acid substituted within the parental AMP, these amino acids are not exactly the same and will thus contribute differently to the physiochemical character of the AMP.

All mutated AMPs total hydrophobic ratios ranged from 39% - 51%, also seen in the parental AMPs with slight percentage changes but still conforming to a range of known AMPs. The total net charge of all the mutated AMPs displayed a range of +2 to +8, except for YYNN4b with a positive net charge of 1, owing to its 0% lysine. All molecular weights remained the same since no amino acids were added, but just one substituted for another one with similar molecular weight.

An instability index of <40 are stable AMPs (providing an estimate of the stability of the AMP in a test tube); a value of <40 was calculated for the mutated AMPs YYNN2a, YYNN4a, YYNN9a, YYNN2b, YYNN9b, YYNN2c, YYNN4c, YYNN8c and YYNN9c. It was observed for AMPs YYNN9a, YYNN9b and YYNN9c, a Boman index less than zero, meaning that they may be good antimicrobial peptides. The mutated anti-*N. meningitidis* AMPs, YYNN2a, YYNN3a, YYNN4a YYNN8a, YYNN2b, YYNN4b, YYNN8b, YYNN2c, YYNN3c, YYNN4c and YYNN8c has a potential to bind to other proteins based on their Boman indices that were less than 2.5 kcal/mol as seen in Table 5.5.

Table 5.5: Physicochemical properties for the derived anti-*N. meningitidis* AMPs

Antimicrobial Peptide Database (APD)									Bactibase		
AMP	Residues	Total hydrophobic ratio	Total net charge	Molecular weight	Protein-binding Potential (Boman index)	Arginine %	Lysine %	Other amino acids with high %	Most Common amino acid	Isoelectric point	Instability Index
YYNN1a	33	39%	5	3842.52	3.34 kcal/mol	21%	3%	Ser, Cys 12%; Leu 9%	Arginine	10.00	75.75
YYNN2a	33	51%	8	3710.572	0.84 kcal/mol	9%	15%	Val 15%; Ile,Ala 12%	Lys,Val	11.79	16.98
YYNN3a	33	45%	2	3874.6	2.16 kcal/mol	15%	3%	Ile,Cys, Ser, Leu 9%	R	8.51	47.36
YYNN4a	34	44%	2	3787.374	2.08 kcal/mol	14%	0%	Leu, Ser 11%	R	8.51	29.08
YYNN5a	33	42%	4	3878.553	3.05 kcal/mol	18%	3%	Ser, Cys 12%; Leu 9%	R	9.11	90.78
YYNN6a	33	42%	4	3775.45	2.87 kcal/mol	18%	3%	Ser, Cys 15%; Leu 9%	R	8.81	84.16
YYNN7a	33	45%	6	3814.577	2.89 kcal/mol	21%	3%	Ser 15%, Cys 12%	R	11.17	70.21
YYNN8a	27	55%	6	3056.792	0.36 kcal/mol	7%	14%	Ala 14%, Val Ile 11%	AK	11.26	45.74
YYNN9a	27	48%	4	2584.138	-0.61 kcal/mol	3%	11%	Gly 25%,Leu 22%	G	11.92	-12.38
AMP	Residues	Total hydrophobic ratio	Total net charge	Molecular weight	Protein-binding Potential (Boman index)	Arginine %	Lysine %	Other amino acids with high %	Most Common amino acid	Isoelectric point	Instability Index
YYNN1b	33	39%	6	3862.553	3.16 kcal/mol	21%	3%	Ser 15%, Cys 12%	R	10.45	59.18
YYNN2b	33	51%	8	3763.635	0.84 kcal/mol	9%	15%	Val 12%; Ile,Ala 12%	K	11.79	36.63
YYNN3b	33	42%	3	3941.694	2.63 kcal/mol	18%	3%	Ile,Cys, Leu 9%	R	8.96	43.32
YYNN4b	34	44%	1	3787.331	2.33 kcal/mol	14%	0%	Ser 11%	R	8.05	44.22
YYNN5b	33	39%	5	3827.501	3.17 kcal/mol	18%	6%	Ser Cys12%	R	9.86	78.33
YYNN6b	33	39%	5	3834.496	2.87 kcal/mol	18%	3%	Ser 15%, Cys 12%	R	9.66	62.73
YYNN7b	33	42%	6	3826.564	2.95 kcal/mol	21%	3%	Ser 15%, Cys Leu 12%	R	11.17	75.06
YYNN8b	27	51%	6	3081.778	0.61 kcal/mol	7%	14%	Ala 14%, Val Ile 11%	AK	11.84	41.76
YYNN9b	27	48%	3	2543.082	-1.1 kcal/mol	0%	11%	Gly 25%,Leu 22%	G	11.1	-15.53
AMP	Residues	Total hydrophobic ratio	Total net charge	Molecular weight	Protein-binding Potential (Boman index)	Arginine %	Lysine %	Other amino acids with high %	Most Common amino acid	Isoelectric point	Instability Index
YYNN1c	33	39%	5	3856.547	3.33 kcal/mol	21%	3%	Ser 15%, Cys Leu 9%	R	11.17	66.16
YYNN2c	33	51%	8	3724.599	0.82 kcal/mol	9%	15%	Val 12%; Ile 15%,Ala 12%	KI	11.79	32.58
YYNN3c	33	45%	2	3914.665	2.19 kcal/mol	15%	3%	Ile,Cys, Ser, Leu 9%	R	8.51	55.77
YYNN4c	34	41%	2	3844.422	2.49 kcal/mol	14%	2%	Ile,Cys, Ser, Leu, Gly, Ala 8%	R	8.51	32.41
YYNN5c	33	42%	4	3892.58	3.03 kcal/mol	18%	3%	Ser Cys Leu 12%	R	9.11	96.62
YYNN6c	33	39%	4	3814.463	3.02 kcal/mol	18%	3%	Ser 15%, Cys 12%	R	9.11	81.88
YYNN7c	33	45%	6	3798.51	3.01 kcal/mol	21%	3%	Ser 15%, Cys 12%	R	11.17	75.06
YYNN8c	27	51%	5	3048.704	0.27 kcal/mol	3%	14%	Ala 14%, Val Ile 11%	AK	11.28	22.7
YYNN9c	27	48%	3	2543.082	-1.1 kcal/mol	0%	11%	Gly 25%,Leu 25%	GL	11.1	-12.38

5.5.4. *De novo* structure prediction of the mutated AMPs

The 3D structures of the mutated AMPs were predicted using the I-TASSER server to see whether the structures of the derived AMPs would change following site directed mutagenesis. The output of I-TASSER server contained statistical indicators used to interpret the results, which provided an estimate of accuracy scoring of the predicted derived AMPs 3-D structure. These statistical indicators are based on the C-score, TM-score, and RMSD.

The results of the predicted 3D structures of the mutated AMPs showed that these peptides have C-score values, which ranged from -2.03 to 1.06. YYNN2a, YYNN2b and YYNN2c were the only mutated AMPs with a score < -1.5. The parental AMP YYNN2 had a C-score of -1.76, which is also smaller than -1.5, meaning that the structure of parental AMP YYNN2 and its derivative AMPs YYNN2a, YYNN2b and YYNN2c was randomly predicted or there was not enough information available for an accurate 3D prediction (Roy *et al.*, 2010). All remaining AMPs represents C-score > -1.5 with a correct fold in structural prediction. All AMPs has TM-scores >0.5, except YYNN2a, YYNN2b and YYNN2c, with TM-scores ranging from 0.47+-0.15, 0.50+-0.15 and 0.49+-0.15, respectively.

The RMSD values with the highest structural requirements suitable for ligand binding were seen in the following AMPs: YYNN3a, YYNN4a, YYNN3b, YYNN4b, YYNN3c and YYNN4c. Medium-resolution models, roughly in the RMSD range of 2–5 Å with a good or reasonable TM-score, resulted in AMPs; YYNN8a, YYNN9a, YYNN8b, YYNN9b, YYNN8c and YYNN9c. As seen in Table 5.6, AMPs with the lowest resolution of <1 Å RMSD , from an otherwise meaningful prediction, i.e. models with an approximately correct topology, TM-score and C-score, were YYNN1a, YYNN5a, YYNN6a, YYNN7a, YYNN1b, YYNN5b, YYNN6b, YYNN7b, YYNN1c, YYNN5c, YYNN6c and YYNN7c, as seen in

Table 5.6. All results of RMSD, TM-score and C-score fall within range of the expected results for correct topology.

Table 5.6: Evaluation scores of the predicted 3D structures by I-TASSER of the derived AMPs

AMP Name and *N. meningitidis* ligand	**C-score**	**Exp.TM-Score**	**Exp.RMSD (Å)**
YYNN1a	0.97	0.85+-0.08	0.5+-0.5
YYNN2a	-2.03	0.47+-0.15	5.8+-3.6
YYNN3a	0.26	0.75+-0.10	1.4+-1.3
YYNN4a	0.55	0.79+-0.09	1.0+-1.0
YYNN5a	0.97	0.85+-0.08	0.5+-0.5
YYNN6a	0.96	0.84+-0.08	0.5+-0.5
YYNN7a	0.91	0.84+-0.08	0.5+-0.5
YYNN8a	-0.56	0.64+-0.13	2.5+-1.9
YYNN9a	-1.21	0.56+-0.15	3.8+-2.6
NhhA	-0.73	0.62+-0.14	9.4+-4.6
AMP Name and *N. meningitidis* ligand	**C-score**	**Exp.TM-Score**	**Exp.RMSD (Å)**
YYNN1b	0.94	0.84+-0.08	0.5+-0.5
YYNN2b	-1.74	0.50+-0.15	5.2+-3.3
YYNN3b	0.12	0.73+-0.11	1.7+-1.5
YYNN4b	0.52	0.78+-0.09	1.1+-1.1
YYNN5b	0.9	0.84+-0.08	0.5+-0.5
YYNN6b	0.94	0.84+-0.08	0.5+-0.5
YYNN7b	1.02	0.85+-0.08	0.5+-0.5
YYNN8b	-0.78	0.61+-0.14	2.9+-2.1
YYNN9b	-1.24	0.56+-0.15	3.8+-2.6
Opc	-0.17	0.69+-0.12	6.3+-3.9
AMP Name and *N. meningitidis* ligand	**C-score**	**Exp.TM-Score**	**Exp.RMSD (Å)**
YYNN1c	0.91	0.85+-0.08	0.5+-0.5
YYNN2c	-1.82	0.49+-0.15	5.4+-3.4
YYNN3c	-0.07	0.70+-0.12	2.0+-1.6
YYNN4c	0.55	0.79+-0.09	1.0+-1.0
YYNN5c	1.06	0.86+-0.07	0.5+-0.5
YYNN6c	0.98	0.85+-0.08	0.5+-0.5
YYNN7c	0.93	0.84+-0.08	0.5+-0.5
YYNN8c	-0.9	0.60+-0.14	3.2+-2.2
YYNN9c	-0.92	0.60+-0.14	3.2+-2.3
PorA	-0.72	0.62+-0.14	8.4+-4.5

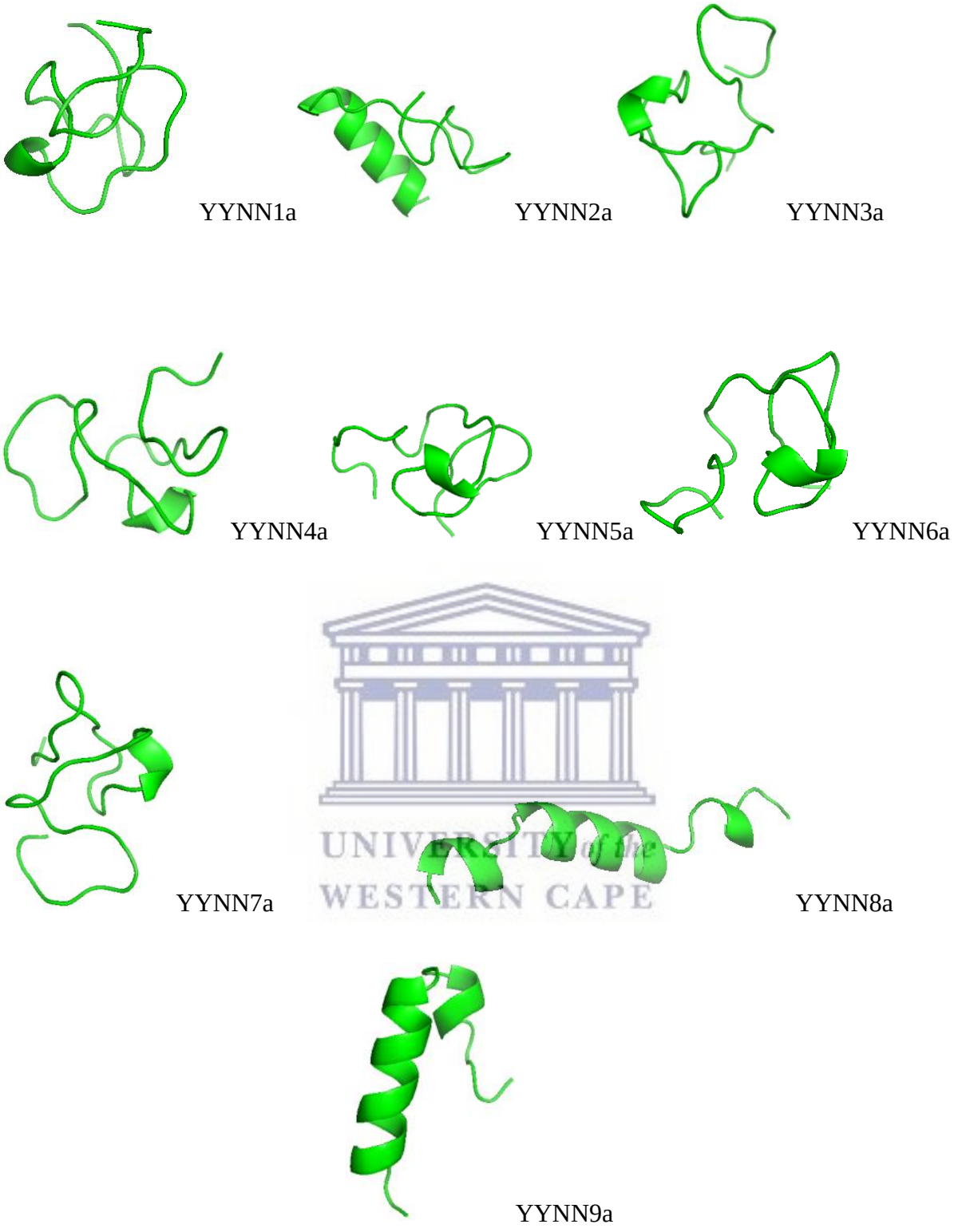


Figure 5.3: Displays the predicted 3D structure of the derived AMPs YYNN1 – 9 using I-TASSER.

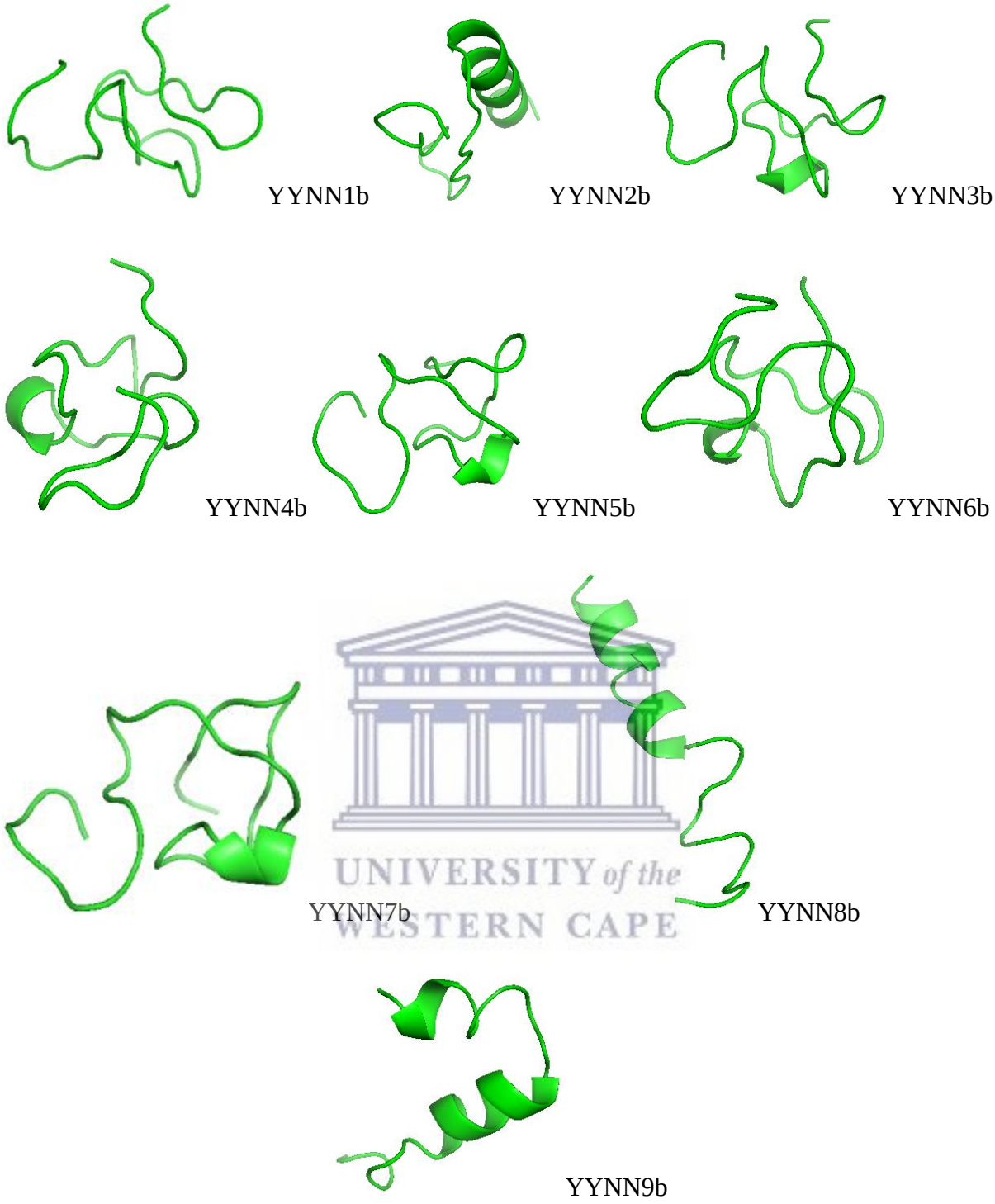


Figure 5.4: Displays the predicted 3-D structure of the derived AMPs YYNN1 - 9 using I-TASSER.

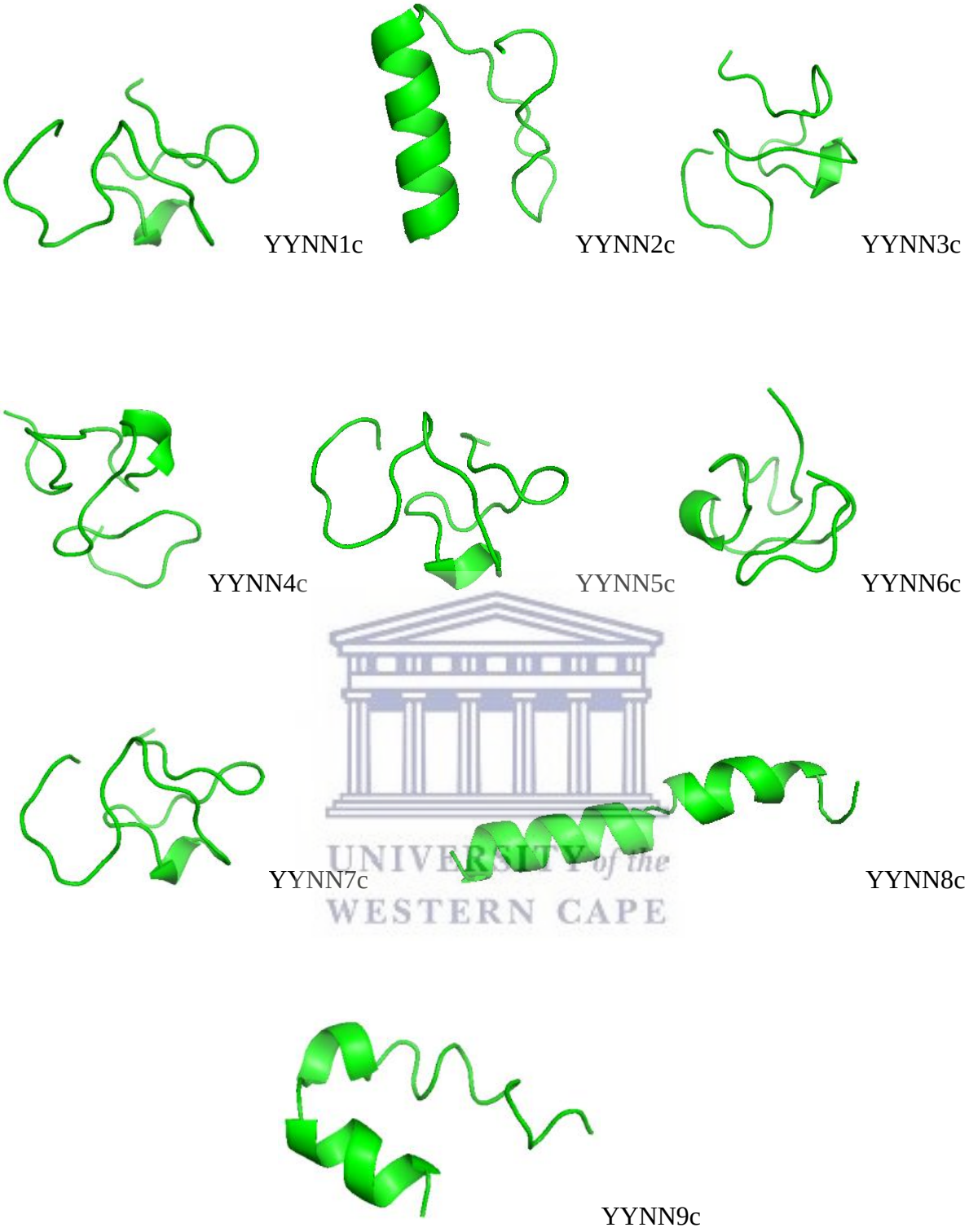


Figure 5.5: Displays the predicted 3-D structure of the derived AMPs YYNN1 - 9 using I-TASSER.

It was observed that the derived AMPS display similar α-helical secondary structures as their parental counterpart AMPs, with slight variation in partial α-helical structure after SDM in certain derived AMPs.

It was observed that the derived YYNN1b is not displaying an extended partial α-helical structure, which was present in the parental YYNN1. The absence of α-helical conformation in YYNN1b was observed which could be justified by the deletion of hydrophilic glutamate amino acid residue and introduction of hydrophobic tyrosine in the derived YYNN1 AMP. The presence of tyrosine plays a role in structure; partially hydrophobic, buried in protein hydrophobic cores and its aromatic side chain can also mean that tyrosine is involved in stacking interactions with other aromatic side-chains and this by its introduction changed the structure of YYNN1.

5.6. Summary

The first aim of this chapter was to use the parental AMPs as templates to generate derivative AMPs that display increased binding affinity for the NhhA, Opc and PorA *N. meningitidis* receptors by identifying 'hotspot' residues responsible for the interaction between the receptors and the parental AMPs using the KFC server. Thereafter, using site directed mutagenesis the ‘non hotspot’ amino acid residues within the binding interaction area were mutated to increase the binding affinity of the AMPs to their interacting proteins (docking studies between the modelled AMP 3D structures and the 3D structures of the *N. meningitidis* receptors shown in chapter 6). Finally, the physicochemical characteristics of each mutated AMP were determined using Bactibase and APD as well as the 3D structures of the mutated AMPs predicted using the algorithm I-TASSER to ensure that the derived AMPs still conform to known AMPs.

The physicochemical properties of the mutated AMPs remained fairly similar following substitution based site-directed mutagenesis and shown based on their physicochemical properties that they still conform to known AMPs.

The results of the predicted 3D structures of the mutated AMPs using I-TASSER showed that these peptides C-score, RMSD-score and TM-score values were within range of the expected results for correct topology.

The mutated AMPs displayed similar α-helical secondary structures as their parental counterpart AMPs, with slight variation in partial α-helical structure. The exception was observed in the derived AMP YYNN1b that did not display an extended partial α-helical structure, which was present in the parental AMP YYNN1.

Chapter 6

In Silico Protein-Peptide Interaction Study

6.1. Introduction

The binding of a ligand to protein receptors underlies a wide variety of recognition processes in biological systems. The understanding of such systems can be enhanced greatly by the development of reliable computational methods to calculate protein–ligand binding constants. Many computational methods using a range of approximations have been developed to estimate both the relative binding affinities of closely related receptors and the absolute binding constants. Widely used simplified methodologies range from docking to continuum electrostatic methods (Woo, 2008). Before a protein-ligand complex is formed, the individual partners that are not a part of hydrophobic surface are involved in hydrogen bonds with the surrounding water. Once the complex is formed, these hydrogen bonds are replaced with hydrogen bonds between the ligand and the protein. The contribution of hydrophobic interactions to protein-ligand binding is normally regarded to be proportional to the size of the hydrophobic surface buried during complex formation. Hydrophobic interactions are also regarded to be the main driving force of conformational change of the receptor upon ligand binding (Yunta, 2016).

Molecular docking is usually performed between a small molecule and a target macromolecule and is defined as the modeling of the 3D structure of a complex from its known constituents, and its combination with a limited amount of NMR- data is extremely powerful and has found a wide range of applications. This is often referred to as ligand–protein docking (Fuentes *et al.*, 2008; Morris and Lim-Wilby, 2008).

Given a protein target, molecular docking generates hundreds of thousands of putative ligand binding orientations/conformations at the active site around the protein. A scoring function is used to rank these ligand orientations/conformations by evaluating the binding tightness of each of the putative complexes. Application of a scoring function is to predict the absolute binding affinity between protein and ligand (Huang *et al.,* 2010).

Experimentally, 3D structures of protein–protein complexes are determined using X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and electron microscopy or a structure obtained by computational techniques (such as homology modeling) (Gromiha *et al.,* 2017; Salmaso and Moro, 2018).

In silico methods for predicting protein–ligand binding sites and protein biochemical functions offers a practical solution to predict the 3D structures of protein–protein complexes using their unbound proteins (Roche *et al.,* 2015; Gromiha *et al.,* 2017).

Small molecules interact with proteins in regions that are accessible and that provide energetically favourable contacts. Geometrically, these binding sites are generally deep, concave shaped regions on the protein surface, referred to alternately as clefts or pockets (Kauffman and Karypis, 2010). The goal of automated molecular docking software is to understand and predict molecular recognition, both structurally, finding likely binding modes, and energetically, predicting binding affinity. Molecular docking helps to provide a 3D structural hypotheses of how a ligand interacts with its target (Morris and Lim-Wilby, 2008).

Protein–peptide docking methods are used most commonly as tools supporting experimental work, for example for the interpretation of ambiguous experimental data, identification of key interactions, or simply for complex visualization (Ciemny *et al.,* 2018).

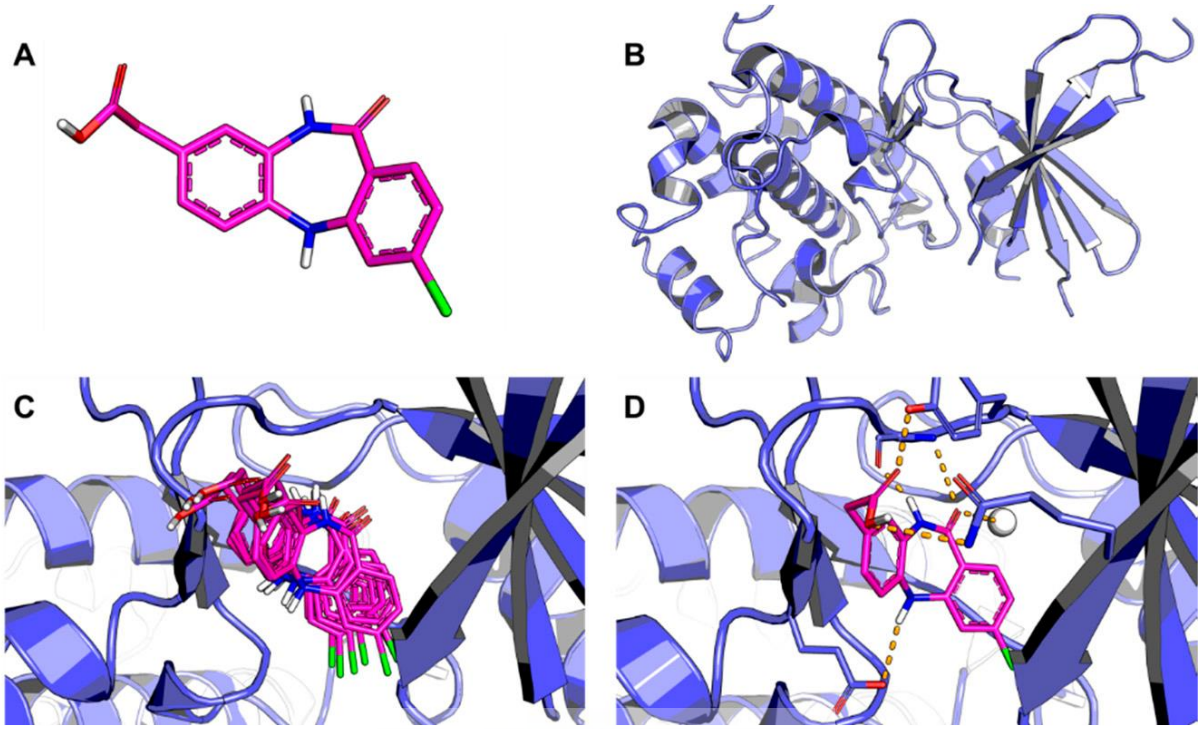


Figure 6.1: Outline of the molecular docking process.

(**A**) 3D structure of the AMP shown as a stick representation; (**B**) 3D structure of the receptor (protein) represented as a cartoon; (**C**) The ligand is docked into the binding cavity of the receptor and the putative conformations are explored; (**D**) The most likely binding conformation and the corresponding intermolecular interactions are identified. The ligand is shown in carbon in magenta and active site residues in carbon in blue. Water is shown as a white sphere and hydrogen bonds are indicated as dashed lines (Ferreira *et al.*, 2015).

6.2. PatchDock

Duhovny *et al.*, 2002 presented an algorithm for unbound (real life) docking of molecules. The high efficiency of the algorithm is the outcome of several factors: (*i*) focusing initial molecular surface fitting on localized, curvature based surface patches; (*ii*) use of Geometric Hashing and Pose Clustering for initial transformation detection; (*iii*) accurate computation of shape complementarity; (*iv*) efficient steric clash detection and geometric fit scoring based on a multi-resolution shape representation and (*v*) utilization of biological information by

focusing on ‘hotspot’ rich surface patches, all contributing to the quality of the results (Duhovny *et al.,* 2002). This algorithm was developed in PatchDock by (Schneidman-Duhovny *et al.,* 2005).

PatchDock Beta 1.3 version is a free online web-server that allows for protein-small ligand molecule docking, available at http://bioinfo3d.cs.tau.ac.il/PatchDock/. The PatchDock method performs structure prediction of protein–protein and protein–small molecule complexes. Within the docking method, the goal is to find the correct association of two interacting molecules given a structural representation for each molecule separately. PatchDock, a very efficient algorithm, is a geometry-based molecular docking algorithm aimed at finding docking transformations that yields good molecular shape complementarity (Schneidman-Duhovny *et al.,* 2005).

The algorithm succeeds in docking of large proteins (antibody with antigen) and small drug molecules. The PatchDock algorithm divides the Connolly dot surface representation of the molecules into concave, convex and flat patches (Schneidman-Duhovny *et al.,* 2005). Then, complementary patches are matched in order to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy. The main reason behind PatchDock’s high efficiency is its fast transformational search, which is driven by local feature matching rather than brute force searching of the six-dimensional transformation space. The run time of PatchDock for two input proteins of average size (about 300 amino acids) is <10 min on a single 1.0 GHz PC processor under the Linux operating system (Schneidman-Duhovny *et al.,* 2005).

6.3. Molecular Graphics Visualisation Tool

6.3.1. RasMol

RasMol is a molecular graphics program intended to visualize proteins, nucleic acids and small molecules for which a 3D structures is available. This standalone software can be downloaded from the RasMol homepage: www.umass.edu/microbio/rasmol. The visualization provides a choice of colour schemes and molecular representation (wireframe, cylinder (Dreiding) stick bonds, alpha carbon trace, space filling (CPK) spheres, macro molecular ribbons (either smooth shaded solid ribbons or parallel strands), hydrogen bonding and dot surface. Additional features such as test labelling for selected atoms, different colour schemes for different parts of the molecule, zoom, rotation, etc. have made this the most popular of all visualization tools (Harisha, 2010).

RasMol can be used as a quick and handy tool for the analysis of biomolecular structures with good results. For many years RasMol is still one of the most used programs for molecular visualization. It is an excellent tool due to its simplicity and its low demand of computer power (Pikora and Gieldon, 2015).

6.4. Aims

This chapter aims to determine the ligand-receptor complex structure with binding affinity scores of the parental and mutated AMPs and the *N. meningitidis* receptors using *in silico* methods. To achieve this, the following objectives will be obtained:

- Complete an *in-silico* protein-peptide interaction of the parental and mutated anti-*N. meningitidis* AMPs and the *N. meningitidis* proteins NhhA, Opc and PorA using PatchDock
- Visualize the *in-silico* binding studies using RasMol

- Analyse the geometric scores of the binding affinity of each anti-*N. meningitidis* AMP (parental and mutated) with each outer membrane protein of *N. meningitidis* (NhhA, Opc and PorA).
- Identify the best candidate AMP (parental or mutated) to be used as a diagnostic molecule based on high binding affinity as well as correct orientation of binding to its receptor.

6.5. Materials and Methods

6.5.1. Molecular Docking Algorithm of PatchDock Based on Shape Complementarity Principles

The nine AMP 3D structures (parental as well as mutated) were docked against each *N. meningitidis* protein (NhhA, Opc and PorA) 3D structure respectively. The binding affinities were recorded and the results were downloaded as PDB files, which represented the structural complex between the AMPs and the *N. meningitidis* proteins. The input was the two molecules in PDB format for each AMP to each protein. The cluster RMSD was set to 4.0 Å and the complex type was selected as "protein-small ligand.

The results received via email included the link to a web page where the predictions are presented. Specific predictions were viewed and the top scoring solutions downloaded as a file in PDB format. The solutions page presents the geometric score (seen in Table 6.1), interface area size and desolvation energy of the 20 top scoring solutions. The geometric score is provided where a higher score is indicative of better binding. PatchDock is one of the only servers that score the binding affinity.

6.5.2. Visualization of the biomolecular 3D structure complex using RasMol

Interaction analysis of the complex formation between the *N. meningitidis* protein and the anti-*N. meningitidis* AMPs were visualized using RasMol 2.7.5.2 software. Each complex

structure is selected and viewed by selecting cartoons for Display and Chain for colour scheme. Output results are seen in Figures 6.2, 6.3 and 6.4.

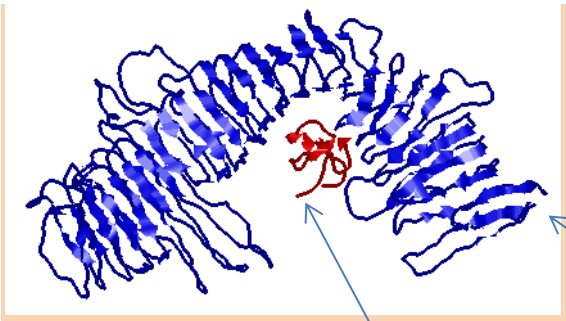
6.6. Results and Discussion

6.6.1. The Protein-Peptide Interaction between the Anti-*N. meningitidis* AMPs and *N. meningitidis* Proteins using PatchDock

The docking results of the structural complex between the AMPs and the proteins were downloaded as PDB files. The objective for studying the structure complex formation between the anti-*N. meningitidis* AMPs and the *N. meningitidis* receptors were not only to predict the interaction of the receptors to the AMPs (ligands), but also to show that the AMPs bind to the receptors at specific sites and with high affinity binding score. The binding of the ligand/s to the receptor/s will facilitate the detection of the *N. meningitidis* bacterium within a patient sample. The results are shown for the C-terminal domain in Figure 6.2, 6.3 and 6.4 in which the receptors (NhhA, Opc and PorA) are seen with a blue surface and the respective ligand (AMPs) are shown in red.

Parental AMPs bound to NhhA

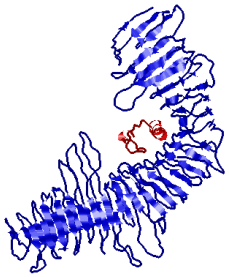
Figure 6.2, displays the interaction of the parental AMPs to the *N. meningitidis* protein NhhA bound at a position between the N-terminus and C-terminus. The C-terminus always faces the periplasmic side of the OM (Hung and Christodoulides, 2013), thus this binding orientation found for all the AMPs, is in the outer membrane of the protein and therefore accessible for detection of *N. meningitidis.*



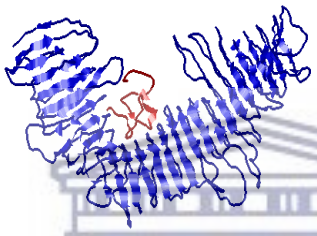
NhhA – Protein (receptor)

YYNN1 – AMP (ligand)

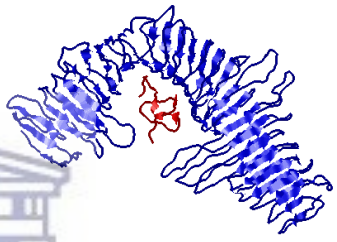
NhhA-YYNN2



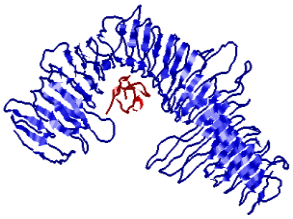
NhhA-YYNN3



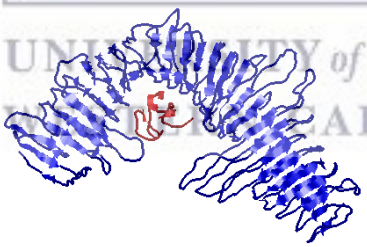
NhhA-YYNN4



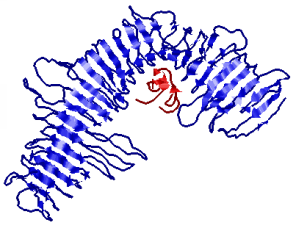
NhhA-YYNN5



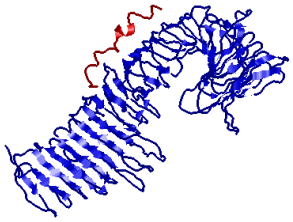
NhhA-YYNN6



NhhA-YYNN7



NhhA-YYNN8



NhhA-YYNN9

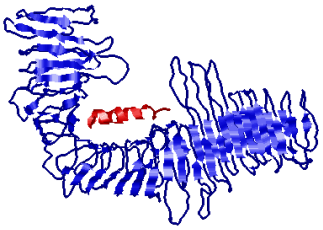
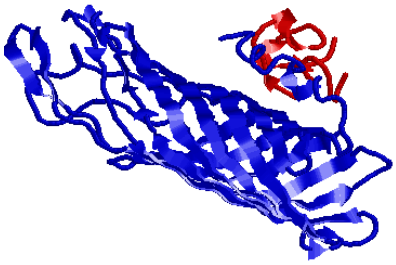


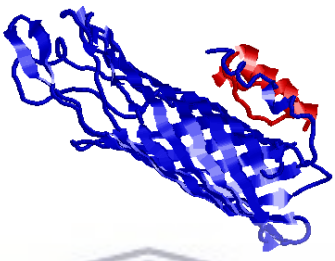
Figure 6.2: Interactions of anti-*N. meningitidis* parental AMPs with NhhA as determined by PatchDock. The red colours depict the anti-*N. meningitidis* AMPs (YYNNs); blue colours represented the NhhA protein

Parental AMPs bound to Opc

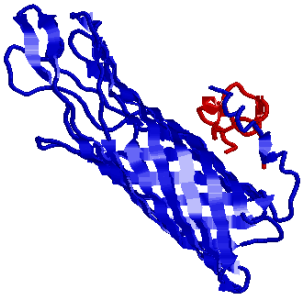
Opc-YYNN1



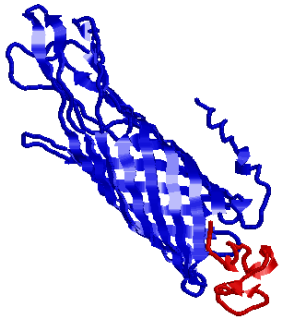
Opc-YYNN2



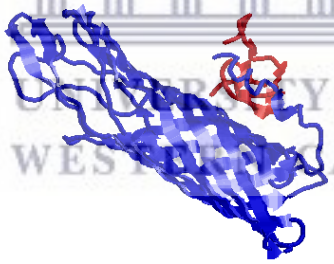
Opc-YYNN3



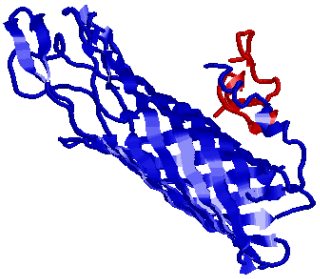
Opc-YYNN4



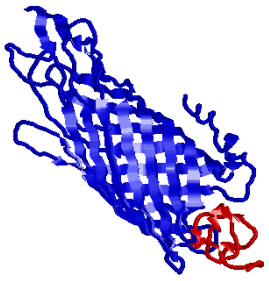
Opc-YYNN5



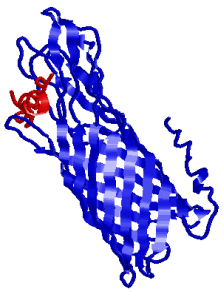
Opc-YYNN6



Opc-YYNN7



Opc-YYNN8



Opc-YYNN9

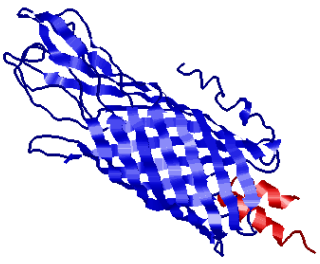


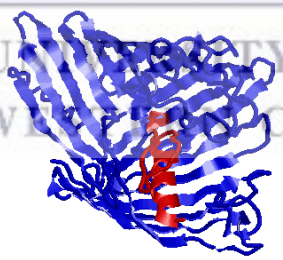
Figure 6.3: Interactions of anti-*N. meningitidis* parental AMPs with Opc as determined by PatchDock. The red colours depicted the anti-*N. meningitidis* AMPs (YYNNs); blue colours represented the Opc protein

As seen in Figure 6.3, YYNN1, YYNN2, YYNN3, YYNN5 and YYNN6 shows binding of the AMPs to the protein Opc at the N-terminal domain which is located in the outer membrane of the *N. meningitidis* bacteria, whereas the AMPs YYNN4, YYNN7 and YYNN9 binds the protein in the periplasmic space of the bacteria and therefore this binding orientation could impede detection of Opc by the AMPs (ligands). YYNN8 binds Opc within the extracellular medium and presented with the highest binding affinity score (as seen in Table 6.1), therefore represents a good candidate for detection of *N. meningitidis*.

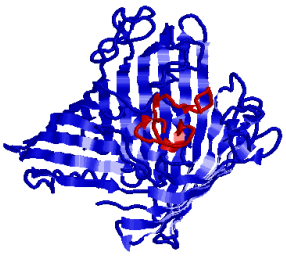
Parental AMPs bound to PorA

PorA-YYNN1

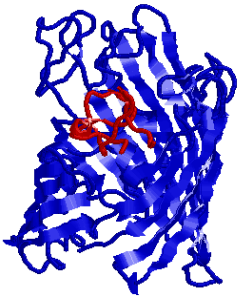
PorA-YYNN2



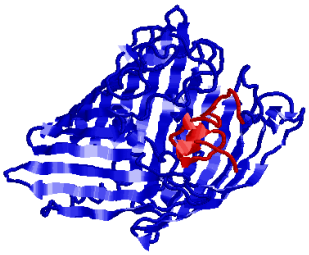
PorA-YYNN3



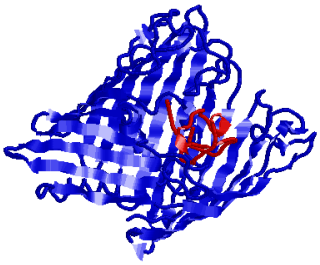
PorA-YYNN4



PorA-YYNN5



PorA-YYNN6



PorA-YYNN7

PorA-YYNN8

PorA-YYNN9

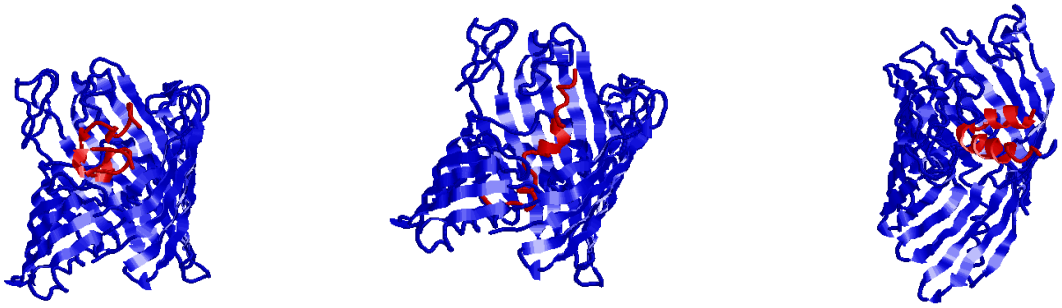


Figure 6.4: Interactions of anti-*N. meningitidis* parental AMPs with PorA as determined by PatchDock. The red colours depicted the anti-*N. meningitidis* AMPs (YYNNs); blue colours represented the PorA protein

All AMPs, as seen in Figure 6.4 are bound to the PorA protein in the outer membrane of the *N. meningitidis* bacteria thus could be used for detection of the bacterium.

6.6.2 Binding affinities

The result output provides the highest geometric score of the complexes formed between proteins NhhA, Opc and PorA and anti-*N. meningitidis* parental AMPs as a PDB file. All the AMPs showed a positive interaction with the three proteins. Highlighted binding affinities represent the highest binding geometric scores of certain AMPs to the respective proteins.

NhhA

Seen in Table 6.1, YYNN5 and YYNN4 showed the highest binding affinities 11904 and 11710, respectively when bound to NhhA, although both peptides did not display the lowest E-value prediction score. A low E-value indicates a prediction of a peptide most likely to be a true *N meningitidis* AMP i.e. a peptide with potentially the highest activity against the bacterium. Since the AMP has to interact with the bacterium it is expected it to have a high binding score. However, this is not necessarily true since AMPs interacts with the bacterial cell wall to exert its activity rather through specific proteins.

AMPs YYNN3, YYNN6 and YYNN9 displayed low binding affinities towards this receptor.

Opc

Opc has the lowest binding affinity scores to all the parental AMPs, with YYNN8 displaying the highest score for this receptor (Table 6.1). This confirms the probability of this peptide to be a true anti-*N. meningitidis* peptide as it also showed the lowest E-value prediction score and fulfils all the physicochemical property requirements of a good AMP.

PorA

The binding affinity scores of individual AMPs to PorA, showed the highest values when compared to the NhhA and Opc proteins docked to all the nine AMPs. All AMPs displayed binding of PorA in the outer membrane and to the 16-strand β-barrel, possibly through the channel pore. Overall, YYNN5 displayed the highest binding affinity geometric score to PorA (Table 6.1). Arginine is the most abundant amino acid within YYNN5 AMP, 18 %, which facilitates binding of AMP to a receptor, with serine and lysine both 12 %. The positive net charge of +4 of YYNN5 is contributed to by the presence of arginine and lysine amino-acids and helps in directing the peptide to the target pathogen through electrostatic attraction. The Boman index value (3.15 kcal/mol) indicates that the peptide YYNN5 is multifunctional with hormone-like activities, although not displaying the lowest E-value prediction scores.

Table 6.1: PatchDock results for each AMP, with the binding affinity geometric Scores

Binding Affinity Geometric scores			
Parental AMPs	**NhhA**	**Opc**	**PorA**
YYNN1	11368	10164	12314
YYNN2	11622	10310	14446
YYNN3	10936	9160	15030
YYNN4	11710	9882	13642
YYNN5	11904	9960	15396
YYNN6	10620	9526	13114
YYNN7	11300	9776	13684
YYNN8	11456	11546	13474
YYNN9	10892	9558	13196

6.6.3. The Protein-Peptide Interaction between the mutated Anti-*N. meningitidis* AMPs and *N. meningitidis* Proteins using PatchDock

The PDB files from the docking study of the mutated AMPs and *N. meningitidis* proteins using PatchDock server were visualized using PyMOL and RasMol software. The analysis of the cartoon representation of *N. meningitidis* protein interaction with the derived AMPs proves that binding occurs at various positions on the respective proteins (Figure 6.5, Figure 6.6. and Figure 6.7) some AMPs as their parental molecules binding to the same position whilst others are showing a shift in binding.

Derived AMPs bound to NhhA

All derivative AMPs displayed binding to NhhA protein at exact same positions as the parental AMPs interaction, except derived YYNN8a has shifted its binding position on the NhhA protein as seen in Figure 6.5.

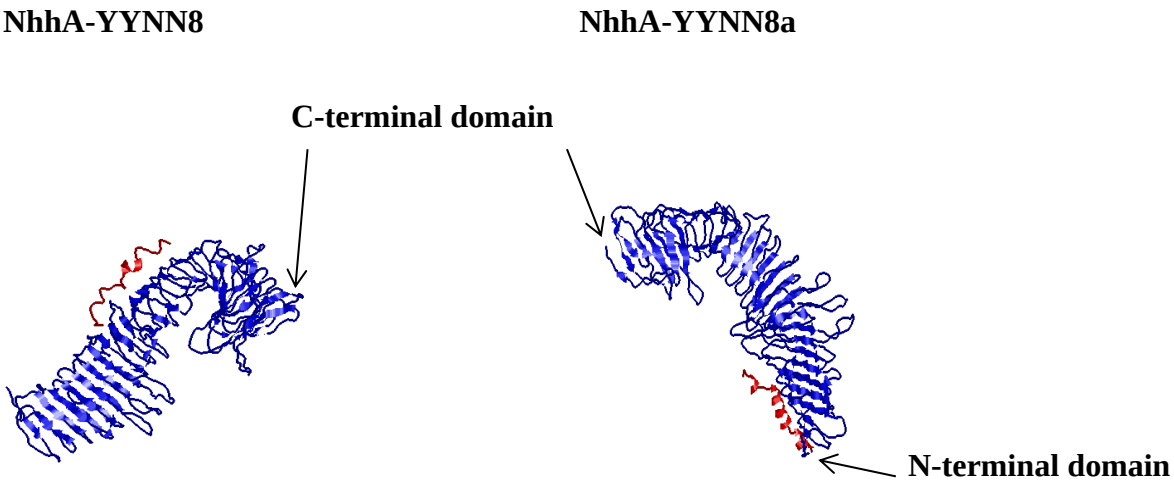


Figure 6.5: The predicted 3D structure of the parental NhhA-YYNN8 and the derived NhhA-YYNN8a complex formation during interaction. The cartoon representation in blue colour is the *N. meningitidis* protein NhhA and the AMPs are represented in red colour.

YYNN8 initially bound at a position between the N-terminal and C-terminal domain, however YYNN8a experienced a noticeable shift in binding, which resulted in its binding at the N-terminal domain (Figure 6.5). The change caused by the introduction of the mutation, displayed a better-predicted fit at the N-terminal domain of the YYNN8 derivative AMP, although showing a decrease in binding affinity from 14456 to 11032. The N-terminal domain presents a larger surface area for binding of AMPs.

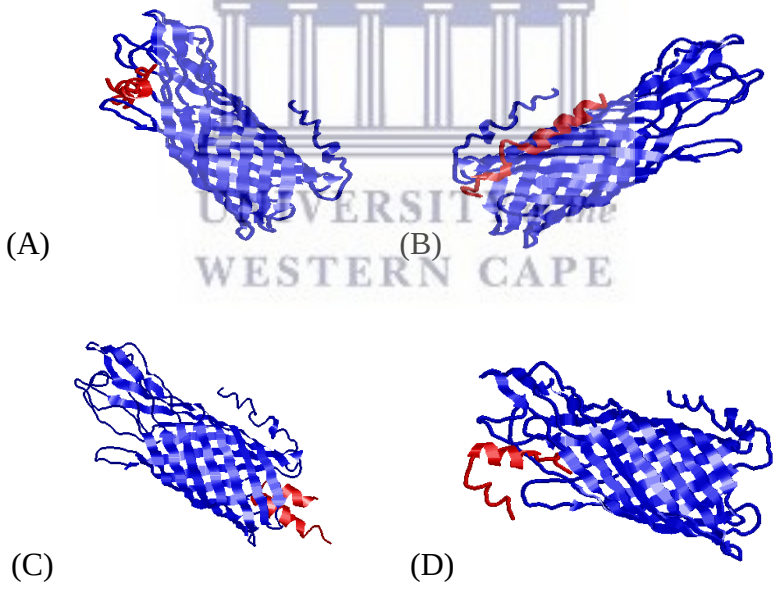
In a study by Williams *et al.*, 2016, several AMPs were identified using HMMER. AMPs were eliminated from further study if it did not bind to HIV p24 protein at the N-terminal as suppose to the C-terminal domain. Several parental AMPs after mutagenesis shift their binding position to the N-terminal domain. The N-terminal binding was preferred since the antibody from the HIV testing kit currently used, competes with p24 antibodies produced by the infected person, leading to a high percentage false negatives (Buttò *et al.*, 2010).

Derived AMPs bound to Opc

parent AMP-YYNN1 derivative AMP-YYNN1b

Figure 6.6: Displays the binding shift of the parent AMP-YYNN1 to the derivative AMP-YYNN1b to Opc. The AMPs are seen in red colour.

YYNN3b and YYNN5b showed the same shift as YYNN1b when binding to Opc.



(A) (B)

(C) (D)

Figure 6.7: The predicted 3D structure of the Opc protein with different AMPs as determined by PatchDock (A) parental AMP YYNN8 bound to Opc (B) derivative AMP YYNN8b bound to Opc, (C) parental AMP YYNN9 bound to Opc and (D) derivative AMP YYNN9b bound to Opc.

The interaction of the Opc protein with the derivative AMPs showed shifting in position of YYNN1b, YYNN3b, YYNN5b, YYNN8b and YYNN9b on this protein (Figure 6.6 and Figure 6.7). This shift displaces the binding of the derivative AMPs YYNN1b, YYNN3b and YYNN5b to the periplasm and therefore the protein would not be accessible by these AMPs for detection of *N. meningitidis.*

In Figure 6.7 C) and D) a clear shift is displayed in binding position of the derived AMP YYNN9b interacting with Opc. This shift presents the AMP YYNN9b bound to Opc in the extracellular medium instead of the periplasmic space, thus being easily available for detection of the bacteria. A shift is also seen in the derived AMP YYNN8b with Opc but, it is still located at the N-terminal domain in the outer membrane.

Derived AMPs bound to PorA

Docking analysis of the PorA-YYNN1c-9c displayed no shifts in positions of the AMPs.

6.6.4 Binding affinities of the mutated AMPs

From the results it was observed that the binding of the derived AMPs to the *N. meningitidis* proteins had both an increase and decrease in binding affinity geometric scores (Table 6.2). The percentage increases of the derived AMPs are discussed further and highlighted in yellow. The binding predictions showed that the binding of certain parental AMPs was higher than their mutated counterparts. As seen in the parental binding analysis, highest binding of YYNN5 to NhhA, YYNN8 to Opc and YYNN5 to PorA, with binding score of 11904, 11546 and 15396, respectively SDM had not resulted in these mutated AMPs to having a higher binding score.

NhhA

AMP YYNN1a (S6T) mutation was made by substituting the hydrophilic serine neutral residue of the parental peptide with threonine amino acid. The introduced amino acid increased the binding score to NhhA receptor by 6.2 % and displayed little difference in physicochemical properties to the parental AMP YYNN1. YYNN3a was mutated by replacing the hydrophobic amino acid proline with a neutral hydrophobic amino acid residue valine (Table 6.2). A percentage increase of the derived AMP YYNN6a to NhhA was observed where hydrophilic glutamine was replaced with cysteine, at position 15 of the parental AMP YYNN6 (Q125C) (Table 6.2). This mutation displays the highest increase of 12.3 % for this receptor.

Opc

Binding is a consequence of net charge and other physiochemical properties such as hydrophilicity or hydrophobicity, size, and functional groups, since these properties are important for protein–protein interactions. The binding affinity of the AMPs was strengthened if the positive charged and/or hydrophobicity percentages of the peptides were increased by an amino acid substitution. After examination of the scores, it can be noted that YYNN9b has the highest binding increase percentage with a net positive percentage of 13.7 % by substituting the hydrophilic serine neutral residue of the parental peptide with threonine amino acid.

An increase in binding affinity was obtained for YYNN3b of 11.8 % when bound to Opc, signifying that the positively charged arginine amino acid substitution played a role in the AMP's increased ability to bind the protein.

PorA

Derivative AMP YYNN2c had an increase in binding affinity from 14446 to 15072 with a net positive percentage of 4.3 %, resulting in this AMP as having the second highest binding affinity score. From the results it was observed that the binding of the derived AMPs to PorA had very high binding scores similar to the parental AMPs but much higher than to the other proteins (Table 6.2).

The binding affinity differences of derived AMPs YYNN1c, YYNN2c, YYNN4c, YYNN6c, YYNN7c and YYNN9c were positive i.e. increased. This increase could be a consequence of the amino acid on the side chain structure change of the AMPs which, played a huge role in the AMP's ability to bind to the protein with an increase in affinity with these AMPs having the required hydrophobicity percentage for strong interaction with the bacterial cell wall

AMP YYNN6c (D17E) mutation was made by substituting the hydrophilic aspartate neutral residue of the parental peptide with glutamate amino acid. The introduced amino acid increased the binding score to PorA receptor by 9.6 % having the highest binding difference and percentage increase compared to the other AMPs binding to PorA and displayed little difference in physicochemical properties to the parental AMP YYNN6.

Table 6.2: Binding affinities, position of parent AMPs and derivative AMPs on *N. meningitidis* proteins. a = binding to NhhA, b = binding to Opc and c = binding to PorA.

Mutated AMPs	Parental AMPs binding score	Derived AMPs binding score	Binding difference	%Increase	Protein	Mutation
YYNN1a	11368	12072	704	6.2%	NhhA	S6T
YYNN2a	11622	11598	-24	-0.2%	NhhA	L26V
YYNN3a	10936	12018	1082	9.9%	NhhA	P9V
YYNN4a	11710	11476	-234	-2.0%	NhhA	E19L
YYNN5a	11904	11626	-278	-2.3%	NhhA	G4V
YYNN6a	10620	11924	1304	12.3%	NhhA	Q15C
YYNN7a	11300	11292	-8	-0.1%	NhhA	T22M
YYNN8a	11456	11032	-424	-3.7%	NhhA	N19C
YYNN9a	10892	10484	-408	-3.7%	NhhA	T26R
YYNN1b	10164	9902	-262	-2.6%	Opc	E19Y
YYNN2b	10310	10238	-72	-0.7%	Opc	F5W
YYNN3b	9160	10240	1080	11.8%	Opc	S18R
YYNN4b	9882	9756	-126	-1.3%	Opc	Y27F
YYNN5b	9960	9168	-792	-8.0%	Opc	H15K
YYNN6b	9526	9156	-370	-3.9%	Opc	E19Y
YYNN7b	9776	10416	640	6.5%	Opc	A12L
YYNN8b	11546	12128	582	5.0%	Opc	N12Q
YYNN9b	9558	10868	1310	13.7%	Opc	S5T
YYNN1c	12314	13268	954	7.7%	PorA	C14M
YYNN2c	14446	15072	626	4.3%	PorA	L26I
YYNN3c	15030	13346	-1684	-11.2%	PorA	G4V
YYNN4c	13642	14404	762	5.6%	PorA	S13K
YYNN5c	15396	12936	-2460	-16.0%	PorA	G11L
YYNN6c	13114	14370	1256	9.6%	PorA	D17E
YYNN7c	13684	14494	810	5.9%	PorA	G4A
YYNN8c	13474	13266	-208	-1.5%	PorA	R2H
YYNN9c	13196	13360	164	1.2%	PorA	V25L

6.7. Summary

In this chapter *in-silico* protein-peptide interaction of the parental and mutated anti-*N. meningitidis* AMPs with the *N. meningitidis* proteins NhhA, Opc and PorA respectively, were accomplished using PatchDock. Visualization of the *in-silico* binding studies was achieved using RasMol. The analysis of the geometric scores of the binding affinity of each anti-*N.*

meningitidis AMP (parental and mutated) with each outer membrane protein of *N. meningitidis* (NhhA, Opc and PorA) was performed to identify the best candidate AMP/s (parental or mutated) to be used as a diagnostic molecule within a Lateral Flow Device (LFD).

The inclusive criteria for that AMP being; having the highest binding affinity as well as correct orientation of binding to its receptor which, will ensure both sensitivity as well as specificity of the LFD for the diagnosis of MD.

YYNN5 showed very high binding affinity geometric scores to the respective *N. meningitidis* proteins with the highest score to PorA. In addition, this AMP was predicted by the HMMER models with a low E-value as well as having all the physicochemical characteristics of an AMP. This AMP also binds to the *N. meningitidis* receptors in the correct binding orientation that will facilitate identification of these receptors within a patient sample. This AMP however, did not show an increase in binding affinity following site directed mutagenesis.

The AMP that showed the most significant increase in binding affinity following site directed mutagenesis was YYNN2c bound to PorA yet, not higher than that observed for parental AMP, YYNN5. Taken together, the results suggest YYNN5 the best candidate to be used in a LFD for the detection of MD followed by YYNN2c or both AMPs to be used to increase the binding affinity to PorA thus increasing the specificity of the identification of *N. meningitidis* within a patient sample.

Chapter 7

General Discussion and Summary

7.1. General discussion

Meningococcal disease (MD) was first described during an outbreak in Geneva in 1805 by Gaspard Vieusseux. In 1884, Italian pathologists Ettore Marchiafava and Angelo Celli described intracellular micrococci in cerebrospinal fluid and in 1887, Anton Wiechselbaum identified *Neisseria meningitidis* (the meningococcus), the causative agent of MD in cerebrospinal fluid and established the connection between the organism and epidemic meningitis (Henry, 2017).

Invasive MD is of major public health importance due to its global distribution, epidemic potential, and predominant disease burden in children, adolescents and fulminant clinical manifestations. Meningococcal meningitis is associated with 5–15% mortality, a rate that has remained relatively unchanged since the 1930s (Pace and Pollard, 2012). Devastating long term sequelae such as amputations, hearing loss and neurodevelopmental disabilities are seen in 11–19% of survivors. Early recognition of MD and its timely treatment is critical in managing and reducing complicated and fatal disease (Pace and Pollard, 2012).

The gold standard for the diagnosis of systemic meningococcal infection is the isolation of *Neisseria meningitidis* (*N. meningitidis)* from culture of a usually sterile body fluid, such as blood or cerebrospinal fluid, or, less commonly, synovial, pleural, or pericardial fluid (Apicella *et al.,* 2009). Although considered the "gold standard" for detection/diagnosis of exposure and infection, culture of live bacterial or viral pathogens is time-consuming, requires significant technical skill and, depending on the agent, biosafety level 3 or 4 is required, thereby necessitating transport to an appropriate laboratory facility (Kulagina *et al.*, 2007).

The frequency of positive blood cultures is 50 to 60 percent, a much lower rate than the frequency of positive Cerebro Spinal Fluid (CSF) cultures (80 to 90 percent), even in patients without overt meningeal symptoms. The TaqMan array card (TAC) is a rapid diagnostic real-time PCR assay that allows simultaneous detection of many viral, bacterial, and parasitic pathogens in blood or cerebrospinal fluid (Apicella *et al.,* 2009). MD has been diagnosed by PCR alone and its sensitivity (reportedly 96%) is not affected by prior administration of antibiotics (Takada *et al.,* 2016). Despite these benefits, PCR has not replaced traditional culture methods for detection of *N. meningitidis* since it cannot be used to determine antimicrobial susceptibility and is not routinely performed by many hospital laboratories. Another limitation is that false-negative results can occur with *N. meningitidis* isolates that possess gene polymorphisms, particularly when a single gene is targeted (Apicella *et al.,* 2009).

Reliable tests for the identification of cases of meningococcal meningitis and serogroup determination are crucial to ensuring proper individual (case-by-case) and collective management of cases and epidemiological surveillance.

Every living organism elaborates antimicrobial peptides (AMPs), as they are vital for survival (Patel and Akhtar, 2017). AMPs are diverse, yet some of them retain high sequence similarity, despite being generated in taxonomically-distant species. Just because the microbial, plant or animal-derived AMPs have shown biological activity, prospecting them as therapeutics for human illnesses appear callous. Additionally, the human immune system is likely to recognize the AMP from other organismal sources as non-self and retaliate by provoking inflammatory responses. Amongst the new-fangled functions of AMPs, their biomarker potential is very promising (Patel and Akhtar, 2017). Peptides are immediate solutions in diagnostic development due to numerous properties such as its small size; therefore rapid and reproducible synthesis, simple and controllable modification, high

stability and non/low toxicity thus overcoming the shortcomings associated with antibodies (Williams *et al.*, 2016).

Schluesener *et al.*, 2012 reports the brain pathology diagnostic marker scope of AMPs as having the potential to be employed in diverse medical fields. Much interest has been focused on development of therapeutic drugs from AMPs, but so far the many promises did not result into any major clinical application (Schluesener *et al.*, 2012). There have been limited reports describing use of AMPs for capture and detection of target analytes i.e. diagnostics.

Combining experimental data with computational biology will ultimately enable better understanding of antimicrobial agent–target interaction and the ability to manipulate biological systems more efficiently. The combination of bioinformatics and relational databases provides the antimicrobial researcher with better tools for analysing, linking and comparing online search results. The development of computational tools depends on knowledge generated from diverse disciplines including mathematics, statistics, computer science, information technology and molecular biology (Hammami and Fliss, 2010). One such tool is the mathematical algorithm, Profile Hidden Markov Models (HMMER), where Tincho *et al.*, 2016, identified putative anti-HIV AMPs using HMMER and Williams *et al.*, 2016, using these identified anti-HIV AMPs to design a LFD prototype accurately detecting both HIV-1 and HIV-2 and providing reproducible results on patient samples with sensitivity, accuracy, low sample requirement, and result interpretation within 15 minutes of the test being taken.

The aim of this research work is to identify novel AMPs with activity against *N. meningitidis* for diagnosis of MD using *in silico* model creation.

7.2. Chapter 2

The aim of this chapter was to construct a sensitive and specific probabilistic model using HIMMER with experimentally validated anti-*N. meningitidis* AMPs as input for identification of putative anti-*N. meningitidis* AMPs from various genome sequences.

Experimentally proven AMPs showing activity towards *N. meningitidis* found in several AMP databases such as APD (Wang and Wang, 2004; Wang *et al.*, 2008; Wang *et al.*, 2016), CAMP (Thomas *et al.*, 2009; Waghu *et al.*, 2014; Waghu *et al.*, 2015), DRAMP (Fan *et al.*, 2016; Liu *et al.*, 2017) and DBAASP (Gogoladze *et al.*, 2014; Pirtskhalava *et al.*, 2015) were retrieved. Data and Literature mining revealed after cross referencing between the databases, 20 AMPs experimentally validated as anti-*N. meningitidis* peptides. A predictive model to identify potentially novel AMPs against the bacteria *N. meningitidis* was created using HMMER. The HMMER model to be used for genome scanning showed more than 95 % confidence to predict a peptide as a putative anti-*N. meningitidis* AMP following calibration of the created model.

Following proteome scanning of the optimized model, several sequences resulted containing both unique/single and multiple domains. Identified sequences containing only unique/single domains were considered, as the single domains had a complete sequence with activity against *Neisseria* and met the cut-off E-value of 0.01.

A final list of nine AMPs using the *in silico* mathematical algorithm HMMER were identified and named YYNN1 – YYNN9 with the smallest E-value seen for the prediction of YYNN8 followed by YYNN9.

Recently, Waghu *et al.*, 2016 studied the use of sequence signatures represented by patterns and HMMs present in experimentally studied AMPs to identify novel AMPs. The study aimed to accelerate the discovery of novel AMPs by exploiting the conserved sequence

signatures present in AMP families. These studies highlight methods such as HMM to predict and design AMPs with great success.

7.3. Chapter 3

The aim of this chapter was to search for *N. meningitidis* receptors and linking it to pathways and secretion mechanisms within the bacterium to determine the most significant outer membrane (OM) protein/s for the novel AMP/s (ligand/s) to interact with. Through the interaction of the AMP/s to selected *N. meningitidis* receptors, the bacterium can be detected within biological samples i.e. a diagnostic test for *N. meningitidis*.

Various search engines such as PubMed/NCBI, Google Scholar and Science Direct were accessed to identify the receptors in the outer membrane of *N. meningitidis* to serve as targets for the identified AMPs. Through an exhaustive literature mining approach, all *N. meningitidis* proteins and their associated pathways and secretion mechanisms were studied to identify the best candidate receptors for the identified AMPs. Three proteins were identified namely NhhA, Opc and PorA based on their selective expression within *N. meningitidis*.

The amino acid sequences of the *N. meningitidis* receptors NhhA, Opc and PorA were extracted using NCBI and UniProt for 3D modelling as well as docking against the 3D models of the identified AMPs (subsequent chapters).

7.4. Chapter 4

The aim of this chapter was to determine whether the peptide sequences identified by HMMER conform to known AMPs both in physicochemical characteristics as well as structure.

The physicochemical properties of the nine identified putative anti-*N. meningitidis* peptide sequences were determined using APD and Bactibase, to ensure that it conformed to known AMPs based on the observed results i.e. their properties. It was ascertained that all the putative anti-*N. meningitidis* peptides have hydrophobic values that are above 30%, which is the anticipated value for hydrophobicity content of an AMP (Table 4.1) with all other physicochemical properties measured also falling within the expected range for known AMPs.

The 3D structure of the *N. meningitidis* receptors and putative AMPs were modelled using the I-TASSER server and visualized using the PyMOL 1.3. Software. From the results AMPs YYNN1, YYNN3 YYNN4, YYNN5, YYNN6 and YYNN7 all have similar structures, represented by an extended partial α-helical structure or loop structure with partial α-helical secondary structure. The AMPs YYNN2, YYNN8 and YYNN9 exhibited α-helical secondary structures. The conformational structures observed for the putative AMPs are also characteristic of known AMPs. Taken together, the peptides identified conform to known AMPs based on physicochemical properties as well as structure. These peptide sequences can thus now be considered *bona vide* AMPs.

Based on the C-score, TM-score and RMSD provided by I-TASSER of the anti-*N. meningitidis* peptides and the *N. meningitidis* protein 3D structures represented models of good prediction and structural topology i.e. correct predicted models. Even though, the structures already exists for the receptor proteins, both here modelled using I-TASSER to ensure uniformity in the criteria used for 3D structure determination.

7.5. Chapter 5

The aim of this chapter was to use the parental AMPs as templates to generate derivative AMPs that display increased predicted binding affinity for the NhhA, Opc and PorA proteins

using site directed mutagenesis. Work within this chapter was carried out using the KFC server to identify "hotspot" amino acid residues within the protein-peptide binding interphase.

After the identification of mutation sensitive amino acids or "hotspot" amino acid residues of the AMPs using the KFC server site directed mutagenesis was carried out by substitution of 'non hotspot' amino acid with amino acids of similar characteristics in an attempt to increase the binding affinity of the AMPs (ligands) to their targets (*N. meningitidis* receptors).

As mentioned above, all amino acids substituted to generate derivate AMPs were of similar characteristics to the amino acids present in the parental AMPs, as to maintain the previously predicted structures of the parental AMPs as well as to still conform to known AMPs following site directed mutagenesis but with increased affinities.

Following mutagenesis, the physicochemical characteristics of each mutated AMP were determined using Bactibase and APD and their 3D structures predicted using I-TASSER to ensure that after SDM the AMPs still retained the properties characteristic of this class of molecules.

The physicochemical properties remained fairly similar following substitution based site-directed mutagenesis using the same parameters for the derived AMPs as what was used for the parental AMPs. Slight changes in physicochemical characteristics post site-directed mutagenesis was observed for certain derived AMPs which, can be explained by the particular amino acid substituted. As indicated, amino acid substitution was carried out with an amino acid of similar properties and not the exact properties as the replaced one.

For structural determination, it was observed that, the derived AMPs display similar α-helical secondary structures as their parental counterparts. The results showed that the derived AMP YYNN1b is not displaying an extended partial α-helical structure, which was present in the

parental YYNN1 which can also be a consequence of the amino acid substituted within the parental AMP.

7.6. Chapter 6

The aim of this chapter was to determine the ligand-receptor complex structure with their associated binding affinity scores of the parental and mutated AMPs with the *N. meningitidis* receptors NhhA, Opc and PorA respectively using PatchDock. Visualisation outputs of the *in silico* binding studies were done using RasMol.

YYNN5 showed the highest binding affinities of all AMPs when bound to PorA and second highest score when bound to NhhA with YYNN8 showing the highest binding score to Opc.

After examination of the scores, it can be noted that YYNN9b had the greatest percentage increase in binding affinity with a net positive percentage of 13.7 %, following site directed mutagenesis, and it still does not have the highest binding score when bound to the *N. meningitidis* proteins. The parental AMP YYNN5 still has the highest binding score bound to PorA protein. Derivative AMP YYNN2c had an increase in binding affinity from 14446 to 15072 with a net positive percentage of 4.3 %, when bound to the PorA protein, resulting in this AMP having the second highest binding affinity score after AMP YYNN5 following mutagenesis. The AMPs also bound the receptors at positions that will allow free binding of ligand and receptor within a biological sample using a lateral flow device (LFD).

The results observed from this chapter as well as previous chapters make YYNN5 the most likely candidate to be a true anti-*N. meningitidis* peptide as it also showed a low E-value prediction score by HMMER (Chapter 2) and fulfils all the physicochemical property requirements of a good AMP (Chapter 4) followed by AMP YYNN2c.

Major outputs from the study

This is the first study to describe the *in silico* creation of a predictive profile for the identification of AMPs to be used as detection molecules for *N. meningitidis*. Similar study was carried out by Tincho *et al.,* 2016, that discovered novel peptides against HIV for the detection of this disease within patient samples.

It is also the first study to aim at discovering peptides for use as diagnostics for *N. meningitidis.* The current detection tests are still based on antibody technology.

Also, there is no current detection test for *N. meningitidis* using lateral flow technology with such a test being created being able to overcome the various shortcoming listed within this work.

Shortcoming of this work

The major shortcoming of this work was the fact that the explored AMP databases had very few sequences for *N. meningitidis.* HMMER requires at the least 100 sequences within the training dataset to create a model with very high discriminatory power. Although, following calibration, the created models performed well in the performance evaluation measures. The output results still need to undergo experimental validation to ensure they possess activity against *N. meningitidis.*

7.7. Future Work

7.7.1. Molecular study

The DNA sequences of the identified bacterial receptors (GenScript, USA) will be cloned into a vector and transformed into *bacterial* cells and purified using the GST purification system. All samples will be analysed by a 12% SDS-PAGE. The purified protein will be lyophilized and stored at -80° C for further use in a binding study.

7.7.2. Peptides synthesis

Selected AMPs will be chemically synthesized by GL Biochem Ltd. using the solid-phase method and will be purified to >98% by reverse-phase High-Pressure Liquid Chromatography and the AMPs will be shipped in a lyophilized form.

7.7.3. Binding studies

Surface Plasmon resonance studies will be carried out between the peptides that showed the highest binding affinities and the *N. meningitidis* proteins it showed the highest binding affinity to, as a confirmation of the observed *in silico* results.

7.7.4. Construction of a lateral flow device

A lateral flow device will be constructed that can detect *N. meningitidis* and tested using patients samples as described by Williams *et al*., 2016 for HIV using AMPs. The best AMPs will be conjugated to gold nanoparticles (AuNPs) and implemented in a Lateral Flow Device (LFD) to determine meningococcal disease using patient sera which could within 15 minutes provide patients with accurate and sensitive diagnosis of meningococcal disease. In the absence of positive patient sera, control sera can be spiked with the various *N. meningitidis* proteins, as to mimic the infection, and applied to the LFD.

Thereafter, a field study will be completed to test n number of patient samples at various stages of infection to assess the consistency and detection power of the LFD prototype.

Table 7.1: Outer Membrane Proteins for the Lateral Flow Device combination test

Opc	Detection of only N. meningitidis within a patient sample
NhhA	Detection of all meningococcal strains within a patient sample
PorA	Detection of all strains of meningococci within a patient sample

Various combinations can be used in the LFD for the detection of *N. meningitidis,* provided the LFD has more than one testing line e.g. T1 and T2, where T1 can detect only the Opc in the patient sample (by using the best AMP) and T2 could detect NhhA (Table 7.1). In this regard, T1 will give a result where it shows *N. meningitidis* is present and potentially virulent due to the presence of Opc and T2 will confirm *N. meningitidis* presence since this OMP (NhhA) is found in all strains (as seen in Figure 7.1.). This being only one combination of AMPs with the potential of more combinations of AMPs introduced to increase sensitivity as well as specificity of the LFD.

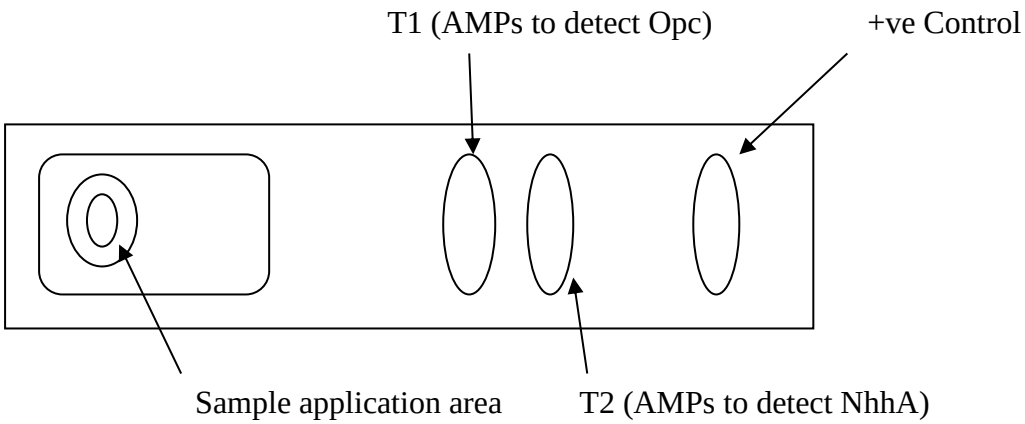


Figure 7.1: A possible Lateral Flow Device combination for *N. meningitidis* detection.

7.7.5. *In silico* work

Using *in silico* analysis of more completely sequenced genomes additional putative AMPs can be identified interact with *N. meningitidis* proteins.

SDM has to be performed on the additional AMPs using positively charged amino acid residues to increase binding scores. Mutations could be carried out making use of a hydrophobic residue such as the amino acid histidine. The end result would be to implement the best suited AMPs in a POC device for detection of MD constantly increasing both sensitivity as well as specificity.

7.8. Outputs

In summary, this work could lead to a new reliable and rapid diagnostic test to detect *N. meningitidis* that should enhance the diagnosis of MD and improve epidemiological surveillance. Rapid testing could help clinicians to identify the many children with MD who are not diagnosed when they first presented to healthcare. This bioinformatic approach potentially identified AMPs with diagnostic potential against *N. meningitidis*. Moreover, the algorithm may be used to prioritize biomarkers in other pathogen species as well.

A POC diagnostic kit of this nature for diagnosis of MD would be easy to access by medical personnel as well as the patients. It will also not require training for personnel to be able to read and understand the result. A POC of this nature will also be very cost effective since it uses AMPs as compared to antibodies which on the average, are produced at a much higher cost than the synthetic process used for AMP synthesis.

The CRDM of NICD also contributes data on numbers and serogroups of *N. meningitidis* and supports diagnostic testing and outbreak response for suspected cases of meningococcal meningitis. One objective for the NICD CRDM is to build local and regional capacity in epidemiology and laboratory diagnostics for respiratory disease and meningitis. Other WHO

meningococcal meningitis program activities include laboratory strengthening to ensure prompt and accurate diagnosis to rapidly confirm the diagnosis of MD. A test of this nature which is much easier to implement compared to the current methods employed for diagnosis of MD will contribute greatly to creating accurate data on this infection to aid International agencies such as WHO to know the exact disease burden as well as predict potential outbreaks in regions with high prevalence of *N. meningitidis* infections.

REFERENCES

a Nijeholt, J. A. L. and A. J. Driessen (2012). "The bacterial Sec-translocase: structure and mechanism." Phil. Trans. R. Soc. B **367**(1592): 1016-1028.

Abby, S. S., J. Cury, J. Guglielmini, B. Néron, M. Touchon and E. P. Rocha (2016). "Identification of protein secretion systems in bacterial genomes." Scientific reports **6**: 23080.

Agnememel, A., F. Traincard, S. Dartevelle, L. Mulard, A. E. Mahamane, O. O. M. Oukem-Boyer, M. Denizon, A. Kacou-N, M. Dosso and B. Gake (2015). "Development and evaluation of a dipstick diagnostic test for Neisseria meningitidis serogroup X." Journal of clinical microbiology **53**(2): 449-454.

Aguilera-Mendoza, L., Y. Marrero-Ponce, R. Tellez-Ibarra, M. T. Llorente-Quesada, J. Salgado, S. J. Barigye and J. Liu (2015). "Overlap and diversity in antimicrobial peptide databases: compiling a non-redundant set of sequences." Bioinformatics **31**(15): 2553-2559.

Ali, O., A. Aseffa, A. B. Omer, T. Lema, T. M. Demissie, Y. Tekletsion, A. Worku, H. G. Xabher, L. Yamuah and R. M. Boukary (2016). "Household transmission of Neisseria meningitidis in the African meningitis belt: a longitudinal cohort study." The Lancet Global Health **4**(12): e989-e995.

Ananiadou, S., D. B. Kell and J.-i. Tsujii (2006). "Text mining and its potential applications in systems biology." Trends in biotechnology **24**(12): 571-579.

Andreae, C. A., R. B. Sessions, M. Virji and D. J. Hill (2018). "Bioinformatic analysis of meningococcal Msf and Opc to inform vaccine antigen design." PloS one **13**(3): e0193940.

Aoki, W., K. Kuroda and M. Ueda (2012). "Next generation of antimicrobial peptides as molecular targeted medicines." Journal of bioscience and bioengineering **114**(4): 365-370.

Apicella, M., B. Calderwood, S. Edvards and R. Thorner (2009). Diagnosis of meningococcal infection, Waltham, MA: UpToDate [online database].

Aruleba, R. T., T. A. Adekiya, B. E. Oyinloye and A. P. Kappo (2018). "Structural Studies of Predicted Ligand Binding Sites and Molecular Docking Analysis of Slc2a4 as a Therapeutic Target for the Treatment of Cancer." International journal of molecular sciences **19**(2): 386.

Attwood, T., A. Gisel, N.-E. Eriksson and E. Bongcam-Rudloff (2011). Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective. Bioinformatics-trends and methodologies, InTech.

Aumentado-Armstrong, T. T., B. Istrate and R. A. Murgita (2015). "Algorithmic approaches to protein-protein interaction site prediction." Algorithms for Molecular Biology **10**(1): 7.

Azad, M. A., H. E. K. Huttunen-Hennelly and C. R. Friedman (2011). "Bioactivity and the First Transmission Electron Microscopy Immunogold Studies of Short De Novo Designed Antimicrobial Peptides." Antimicrobial agents and chemotherapy.

Bahar, A. A. and D. Ren (2013). "Antimicrobial peptides." Pharmaceuticals **6**(12): 1543-1575.

Bahr, N. C. and D. R. Boulware (2014). "Methods of rapid diagnosis for the etiology of meningitis in adults." Biomarkers in medicine **8**(9): 1085-1103.

Batista, R. S., A. P. Gomes, J. L. D. Gazineo, P. S. B. Miguel, L. A. Santana, L. Oliveira and M. Geller (2017). "Meningococcal disease, a clinical and epidemiological review." Asian Pacific Journal of Tropical Medicine.

Baxevanis, A. D. and A. Bateman (2015). "The importance of biological databases in biological discovery." Current protocols in bioinformatics **50**(1): 1.1. 1-1.1. 8.

Bechinger, B. and S.-U. Gorr (2017). "Antimicrobial peptides: mechanisms of action and resistance." Journal of dental research **96**(3): 254-260.

Beran, O., D. Lawrence, N. Andersen, O. Dzupova, J. Kalmusova, M. Musilek and M. Holub (2009). "Sequential analysis of biomarkers in cerebrospinal fluid and serum during invasive meningococcal disease." European journal of clinical microbiology & infectious diseases **28**(7): 793-799.

Bernstein, H. D. (2015). "Looks can be deceiving: recent insights into the mechanism of protein secretion by the autotransporter pathway." Molecular microbiology **97**(2): 205-215.

Betts, M. J. and R. B. Russell (2007). "Amino-acid properties and consequences of substitutions." Bioinformatics for geneticists **2**: 311-339.

Boman, H. (2003). "Antibacterial peptides: basic facts and emerging concepts." Journal of internal medicine **254**(3): 197-215.

Borrow, R., P. Alarcón, J. Carlos, D. A. Caugant, H. Christensen, R. Debbag, P. De Wals, G. Echániz-Aviles, J. Findlow and C. Head (2017). "The Global Meningococcal Initiative: global epidemiology, the impact of vaccines on meningococcal disease and the importance of herd protection." Expert review of vaccines **16**(4): 313-328.

Boughorbel, S., F. Jarray and M. El-Anbari (2017). "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric." PloS one **12**(6): e0177678.

Bourke, T. W., D. J. Fairley and M. D. Shields (2010). "Rapid diagnosis of meningococcal disease." Expert review of anti-infective therapy **8**(12): 1321-1323.

Boutet, E., D. Lieberherr, M. Tognolli, M. Schneider, P. Bansal, A. J. Bridge, S. Poux, L. Bougueleret and I. Xenarios (2016). UniProtKB/Swiss-Prot, the manually annotated section

of the UniProt KnowledgeBase: how to use the entry view. Plant Bioinformatics, Springer**:** 23-54.

Brandtzaeg, P., A. Bjerre, R. Øvstebø, B. Brusletto, G. B. Joø and P. Kierulf (2001). "Invited review: Neisseria meningitidis lipopolysaccharides in human pathology." Journal of endotoxin research **7**(6): 401-420.

Brandtzaeg, P. and M. van Deuren (2012). "Neisseria meningitidis: advanced methods and protocols."

Brouwer, M. C., A. R. Tunkel and D. van de Beek (2010). "Epidemiology, diagnosis, and antimicrobial treatment of acute bacterial meningitis." Clinical microbiology reviews **23**(3): 467-492.

Bryant, P. A., H. Y. Li, A. Zaia, J. Griffith, G. Hogg, N. Curtis and J. R. Carapetis (2004). "Prospective study of a real-time PCR that is highly sensitive, specific, and clinically useful for diagnosis of meningococcal disease in children." Journal of clinical microbiology **42**(7): 2919-2925.

Bullard, R. S., W. Gibson, S. K. Bose, J. K. Belgrave, A. C. Eaddy, C. J. Wright, D. J. Hazen-Martin, J. M. Lage, T. E. Keane and T. A. Ganz (2008). "Functional analysis of the host defense peptide Human Beta Defensin-1: new insight into its potential role in cancer." Molecular immunology **45**(3): 839-848.

Buttò, S., B. Suligoi, E. Fanales-Belasio and M. Raimondo (2010). "Laboratory diagnostics for HIV infection." Annali dell'Istituto superiore di sanita **46**: 24-33.

Carbonnelle, E., D. J. Hill, P. Morand, N. J. Griffiths, S. Bourdoulous, I. Murillo, X. Nassif and M. Virji (2009). "Meningococcal interactions with the host." Vaccine **27**: B78-B89.

Carrigan, P. E., P. Ballar and S. Tuzmen (2011). Site-directed mutagenesis. Disease Gene Identification, Springer**:** 107-124.

Chang, Q., Y.-L. Tzeng and D. S. Stephens (2012). "Meningococcal disease: changes in epidemiology and prevention." Clinical epidemiology **4**: 237.

Chanteau, S., S. Dartevelle, A. E. Mahamane, S. Djibo, P. Boisier and F. Nato (2006). "New rapid diagnostic tests for Neisseria meningitidis serogroups A, W135, C, and Y." PLoS medicine **3**(9): e337.

Chen, P., J. Li, L. Wong, H. Kuwahara, J. Z. Huang and X. Gao (2013). "Accurate prediction of hot spot residues through physicochemical characteristics of amino acid sequences." Proteins: Structure, Function, and Bioinformatics **81**(8): 1351-1362.

Chicco, D. (2017). "Ten quick tips for machine learning in computational biology." BioData mining **10**(1): 35.

Cho, K.-i., D. Kim and D. Lee (2009). "A feature-based approach to modeling protein–protein interaction hot spots." Nucleic acids research **37**(8): 2672-2687.

Chow, J., K. Uadiale, A. Bestman, C. Kamau, D. Caugant, A. Shehu and J. Greig (2016). "Invasive Meningococcal Meningitis Serogroup C Outbreak in Northwest Nigeria, 2015-Third Consecutive Outbreak of a New Strain." PLoS currents **8**.

Ciemny, M., M. Kurcinski, K. Kamel, A. Kolinski, N. Alam, O. Schueler-Furman and S. Kmiecik (2018). "Protein–peptide docking: opportunities and challenges." Drug discovery today.

Cohen, C., E. Singh, H. M. Wu, S. Martin, L. de Gouveia, K. P. Klugman, S. Meiring, N. Govender, A. von Gottberg, R. Group for Enteric and M. d. S. i. S. Africa (2010). "Increased

incidence of meningococcal disease in HIV-infected individuals associated with higher case-fatality ratios in South Africa." Aids **24**(9): 1351-1360.

Cohen, J. (2004). "Bioinformatics—an introduction for computer scientists." ACM Computing Surveys (CSUR) **36**(2): 122-158.

Colquitt, R. B., D. A. Colquhoun and R. H. Thiele (2011). "In silico modelling of physiologic systems." Best practice & research Clinical anaesthesiology **25**(4): 499-510.

Coordinators, N. R. (2017). "Database resources of the national center for biotechnology information." Nucleic acids research **45**(Database issue): D12.

Coureuil, M., O. Join-Lambert, H. Lécuyer, S. Bourdoulous, S. Marullo and X. Nassif (2012). "Mechanism of meningeal invasion by Neisseria meningitidis." Virulence **3**(2): 164-172.

Cukuroglu, E., H. B. Engin, A. Gursoy and O. Keskin (2014). "Hot spots in protein–protein interfaces: Towards drug discovery." Progress in biophysics and molecular biology **116**(2-3): 165-173.

Cunha, C. S. E., N. J. Griffiths and M. Virji (2010). "Neisseria meningitidis Opc invasin binds to the sulphated tyrosines of activated vitronectin to attach to and invade human brain endothelial cells." PLos pathogens **6**(5): e1000911.

da Cunha, N. B., N. B. Cobacho, J. F. Viana, L. A. Lima, K. B. Sampaio, S. S. Dohms, A. C. Ferreira, C. de la Fuente-Núñez, F. F. Costa and O. L. Franco (2017). "The next generation of antimicrobial peptides (AMPs) as molecular therapeutic tools for the treatment of diseases with social and economic impacts." Drug discovery today **22**(2): 234-248.

da Silva, F. P. and M. C. C. Machado (2012). "Antimicrobial peptides: clinical relevance and therapeutic implications." Peptides **36**(2): 308-314.

Dagliyan, O., E. A. Proctor, K. M. D'Auria, F. Ding and N. V. Dokholyan (2011). "Structural and dynamic determinants of protein-peptide recognition." Structure **19**(12): 1837-1845.

Darnell, S. J., L. LeGault and J. C. Mitchell (2008). "KFC Server: interactive forecasting of protein interaction hot spots." Nucleic acids research **36**(suppl_2): W265-W269.

Darnell, S. J., D. Page and J. C. Mitchell (2007). "An automated decision-tree approach to predicting protein interaction hot spots." Proteins: Structure, Function, and Bioinformatics **68**(4): 813-823.

DeLano, W. L. (2002). "Pymol: An open-source molecular graphics tool." CCP4 Newsletter On Protein Crystallography **40**: 82-92.

DeLano, W. L. and S. Bromberg (2004). "PyMOL user's guide."

Doering, J. A., S. Lee, K. Kristiansen, L. Evenseth, M. G. Barron, I. Sylte and C. A. LaLone (2018). "In silico site-directed mutagenesis informs species-specific predictions of chemical susceptibility derived from the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool." Toxicological Sciences.

Doran, K. S., M. Fulde, N. Gratz, B. J. Kim, R. Nau, N. Prasadarao, A. Schubert-Unkmeir, E. I. Tuomanen and P. Valentin-Weigand (2016). "Host–pathogen interactions in bacterial meningitis." Acta neuropathologica **131**(2): 185-209.

du Plessis, M., C. Moodley, K. M. Mothibeli, A. Fali, K. P. Klugman and A. Von Gottberg (2012). "Population snapshot of invasive serogroup B meningococci in South Africa, 2005-2008." Journal of clinical microbiology: JCM. 00401-00412.

Duhovny, D., R. Nussinov and H. J. Wolfson (2002). Efficient unbound docking of rigid molecules. International workshop on algorithms in bioinformatics, Springer.

Dwilow, R. and S. Fanella (2015). "Invasive Meningococcal Disease in the 21st Century—An Update for the Clinician." Current Neurology and Neuroscience Reports **15**(3): 2.

Ebenhan, T., O. Gheysens, H. G. Kruger, J. R. Zeevaart and M. M. Sathekge (2014). "Antimicrobial peptides: their role as infection-selective tracers for molecular imaging." BioMed research international **2014**.

Eddy, S. (2003). "HMMER user's guide. biological sequence analysis using profile hidden Markov models."

Eddy, S. R. (1998). "Profile hidden Markov models." Bioinformatics (Oxford, England) **14**(9): 755-763.

Eddy, S. R. (2011). "Accelerated profile HMM searches." PLoS computational biology **7**(10): e1002195.

El Nahas, M., S. Kassim and N. Shikoun (2012). "profile hidden markov model for detection and prediction of hepatitis C virus mutation."

Epand, R. F., G. Wang, B. Berno and R. M. Epand (2009). "Lipid segregation explains selective toxicity of a series of fragments derived from the human cathelicidin LL-37." Antimicrobial agents and chemotherapy **53**(9): 3705-3714.

Epand, R. M. (2007). "Detecting the presence of membrane domains using DSC." Biophysical chemistry **126**(1-3): 197-200.

Epand, R. M. and R. F. Epand (2011). "Bacterial membrane lipids in the action of antimicrobial agents." Journal of Peptide Science **17**(5): 298-305.

Epand, R. M., S. Rotem, A. Mor, B. Berno and R. F. Epand (2008). "Bacterial membranes as predictors of antimicrobial potency." Journal of the American Chemical Society **130**(43): 14346-14352.

Estabrook, M. M., J. M. Griffiss and G. A. Jarvis (1997). "Sialylation of Neisseria meningitidis lipooligosaccharide inhibits serum bactericidal activity by masking lacto-N-neotetraose." Infection and immunity **65**(11): 4436-4444.

Falagas, M. E., E. I. Pitsouni, G. A. Malietzis and G. Pappas (2008). "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses." The FASEB journal **22**(2): 338-342.

Fan, L., J. Sun, M. Zhou, J. Zhou, X. Lao, H. Zheng and H. Xu (2016). "DRAMP: a comprehensive data repository of antimicrobial peptides." Scientific reports **6**: 24482.

Ferreira, L. G., R. N. dos Santos, G. Oliva and A. D. Andricopulo (2015). "Molecular docking and structure-based drug design strategies." Molecules **20**(7): 13384-13421.

Finn, R. D., J. Clements, W. Arndt, B. L. Miller, T. J. Wheeler, F. Schreiber, A. Bateman and S. R. Eddy (2015). "HMMER web server: 2015 update." Nucleic acids research **43**(W1): W30-W38.

Finn, R. D., J. Clements and S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." Nucleic acids research **39**(suppl_2): W29-W37.

Fjell, C. D., J. A. Hiss, R. E. Hancock and G. Schneider (2012). "Designing antimicrobial peptides: form follows function." Nature reviews Drug discovery **11**(1): 37.

Fjell, C. D., H. Jenssen, P. Fries, P. Aich, P. Griebel, K. Hilpert, R. E. Hancock and A. Cherkasov (2008). "Identification of novel host defense peptides and the absence of α-defensins in the bovine genome." Proteins: Structure, Function, and Bioinformatics **73**(2): 420-430.

Freeman, M. K., S. A. Lauderdale, M. G. Kendrach and T. W. Woolley (2009). "Google Scholar versus PubMed in locating primary literature to answer drug-related questions." Annals of Pharmacotherapy **43**(3): 478-484.

Frosch, M. and M. C. Maiden (2006). Handbook of meningococcal disease: infection biology, vaccination, clinical management, John Wiley & Sons.

Fuentes, G., A. D. van Dijk and A. M. Bonvin (2008). Nuclear magnetic resonance-based modeling and refinement of protein three-dimensional structures and their complexes. Molecular Modeling of Proteins, Springer**:** 229-255.

Gabere, M. N. and W. S. Noble (2017). "Empirical comparison of web-based antimicrobial peptide prediction tools." Bioinformatics **33**(13): 1921-1929.

Gabutti, G., A. Stefanati and P. Kuhdari (2015). "Epidemiology of Neisseria meningitidis infections: case distribution by age and relevance of carriage." Journal of preventive medicine and hygiene **56**(3): E116.

Gasparini, R., D. Amicizia, P. L. Lai and D. Panatto (2012). "Neisseria meningitidis: pathogenetic mechanisms to overcome the human immune defences." Journal of preventive medicine and hygiene **53**(2).

Gasparini, R., D. Panatto, N. Bragazzi, P. Lai, A. Bechini, M. Levi, P. Durando and D. Amicizia (2015). "How the knowledge of interactions between meningococcus and the human immune system has been used to prepare effective Neisseria meningitidis vaccines." Journal of immunology research **2015**.

Gianchecchi, E., A. Torelli, G. Piccini, S. Piccirella and E. Montomoli (2015). "Neisseria meningitidis infection: who, when and where?" Expert review of anti-infective therapy **13**(10): 1249-1263.

Gioia, C. A. C., A. P. S. de Lemos, M. C. O. Gorla, R. A. Mendoza-Sassi, T. Ballester, A. Von Groll, B. Wedig, N. de Vargas Ethur, L. Bragança and L. G. Milagres (2015). "Detection of Neisseria meningitidis in asymptomatic carriers in a university hospital from Brazil." Revista Argentina de microbiologia **47**(4): 322-327.

Giuliani, A., G. Pirri and S. Nicoletto (2007). "Antimicrobial peptides: an overview of a promising class of therapeutics." Open Life Sciences **2**(1): 1-33.

Gogoladze, G., M. Grigolava, B. Vishnepolsky, M. Chubinidze, P. Duroux, M.-P. Lefranc and M. Pirtskhalava (2014). "DBAASP: database of antimicrobial activity and structure of peptides." FEMS microbiology letters **357**(1): 63-68.

Gowin, E., J. Wysocki, D. Avonts, D. Januszkiewicz-Lewandowska and M. Michalak (2016). "Usefulness of inflammatory biomarkers in discriminating between bacterial and aseptic meningitis in hospitalized children from a population with low vaccination coverage." Archives of medical science: AMS **12**(2): 408.

Green, E. R. and J. Mecsas (2016). "Bacterial secretion systems–an overview." Microbiology spectrum **4**(1).

Grijpstra, J., J. Arenas, L. Rutten and J. Tommassen (2013). "Autotransporter secretion: varying on a theme." Research in microbiology **164**(6): 562-582.

Grinter, S. Z. and X. Zou (2014). "Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design." Molecules **19**(7): 10150-10176.

Gromiha, M. M., K. Yugandhar and S. Jemimah (2017). "Protein–protein interactions: scoring schemes and binding affinity." Current opinion in structural biology **44**: 31-38.

Guaní-Guerra, E., T. Santos-Mendoza, S. O. Lugo-Reyes and L. M. Terán (2010). "Antimicrobial peptides: general overview and clinical implications in human health and disease." Clinical immunology **135**(1): 1-11.

Hammami, R. and I. Fliss (2010). "Current trends in antimicrobial agent research: chemo-and bioinformatics approaches." Drug discovery today **15**(13-14): 540-546.

Hammami, R., A. Zouhir, J. B. Hamida and I. Fliss (2007). "BACTIBASE: a new web-accessible database for bacteriocin characterization." Bmc Microbiology **7**(1): 89.

Hammami, R., A. Zouhir, C. Le Lay, J. B. Hamida and I. Fliss (2010). "BACTIBASE second release: a database and tool platform for bacteriocin characterization." Bmc Microbiology **10**(1): 22.

Hancock, R. E. and G. Diamond (2000). "The role of cationic antimicrobial peptides in innate host defences." Trends in microbiology **8**(9): 402-410.

Harisha, S. (2010). Fundamentals of Bioinformatics, IK International Pvt Ltd.

Harrison, L. H., M. A. Pass, A. B. Mendelsohn, M. Egri, N. E. Rosenstein, A. Bustamante, J. Razeq and J. C. Roche (2001). "Invasive meningococcal disease in adolescents and young adults." Jama **286**(6): 694-699.

Harrison, L. H., S. I. Pelton, A. Wilder-Smith, J. Holst, M. A. Safadi, J. A. Vazquez, M.-K. Taha, F. M. LaForce, A. Von Gottberg and R. Borrow (2011). "The Global Meningococcal Initiative: recommendations for reducing the global burden of meningococcal disease." Vaccine **29**(18): 3363-3371.

Harrison, L. H., C. L. Trotter and M. E. Ramsay (2009). "Global epidemiology of meningococcal disease." Vaccine **27**: B51-B63.

Harrison, O. B., H. Claus, Y. Jiang, J. S. Bennett, H. B. Bratcher, K. A. Jolley, C. Corton, R. Care, J. T. Poolman and W. D. Zollinger (2013). "Description and nomenclature of Neisseria meningitidis capsule locus." Emerging infectious diseases **19**(4): 566.

Henry, R. (2017). "Etymologia: Meningococcal Disease." Emerging Infectious Diseases **23**(7): 1187.

Herwald, H. and A. Egesten (2011). Sepsis-Pro-Inflammatory and Anti-Inflammatory Responses, Karger Medical and Scientific Publishers.

Hill, D. J., N. J. Griffiths, E. Borodina, C. A. Andreae, R. B. Sessions and M. Virji (2015). "Identification and therapeutic potential of a vitronectin binding region of meningococcal msf." PLos one **10**(3): e0124133.

Hill, D. J., N. J. Griffiths, E. Borodina and M. Virji (2010). "Cellular and molecular biology of Neisseria meningitidis colonization and invasive disease." Clinical science **118**(9): 547-564.

Hill, D. J. and M. Virji (2012). Meningococcal receptors and molecular targets of the host. Neisseria meningitidis, Springer**:** 143-152.

Hogeweg, P. (2011). "The roots of bioinformatics in theoretical biology." PLoS computational biology **7**(3): e1002021.

Holland, I. B. (2010). The extraordinary diversity of bacterial protein secretion mechanisms. Protein Secretion, Springer**:** 1-20.

Hooda, Y., H. E. Shin, T. J. Bateman and T. F. Moraes (2017). "Neisserial surface lipoproteins: structure, function and biogenesis." Pathogens and disease **75**(2).

Hoyos-Nogués, M., F. Gil and C. Mas-Moruno (2018). "Antimicrobial Peptides: Powerful Biorecognition Elements to Detect Bacteria in Biosensing Technologies." Molecules **23**(7): 1683.

Hu, S.-S., P. Chen, B. Wang and J. Li (2017). "Protein binding hot spots prediction from sequence only by a new ensemble learning method." Amino acids **49**(10): 1773-1785.

Huang, C.-C. and Z. Lu (2015). "Community challenges in biomedical text mining over 10 years: success, failure and the future." Briefings in bioinformatics **17**(1): 132-144.

Huang, S.-Y., S. Z. Grinter and X. Zou (2010). "Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions." Physical Chemistry Chemical Physics **12**(40): 12899-12908.

Huang, S.-Y. and X. Zou (2010). "Advances and challenges in protein-ligand docking." International journal of molecular sciences **11**(8): 3016-3034.

Huang, Y., J. Huang and Y. Chen (2010). "Alpha-helical cationic antimicrobial peptides: relationships of structure and function." Protein & cell **1**(2): 143-152.

Hung, M.-C. and M. Christodoulides (2013). "The biology of Neisseria adhesins." Biology **2**(3): 1054-1109.

Hung, M.-C., J. E. Heckels and M. Christodoulides (2013). "The adhesin complex protein (ACP) of Neisseria meningitidis is a new adhesin with vaccine potential." MBio **4**(2): e00041-00013.

Ioannou, Z.-M., C. Makris, G. P. Patrinos and G. Tzimas (2014). "A set of novel mining tools for efficient biological knowledge discovery." Artificial Intelligence Review **42**(3): 461-478.

Izadpanah, A. and R. L. Gallo (2005). "Antimicrobial peptides." Journal of the American Academy of Dermatology **52**(3): 381-390.

Jacob, E. and R. Unger (2007). "A tale of two tails: why are terminal residues of proteins exposed?" Bioinformatics **23**(2): e225-e230.

Jacob-Dubuisson, F., J. Guérin, S. Baelen and B. Clantin (2013). "Two-partner secretion: as simple as it sounds?" Research in microbiology **164**(6): 583-595.

Jafri, R. Z., A. Ali, N. E. Messonnier, C. Tevi-Benissan, D. Durrheim, J. Eskola, F. Fermon, K. P. Klugman, M. Ramsay and S. Sow (2013). "Global epidemiology of invasive meningococcal disease." Population health metrics **11**(1): 17.

Jensen, L. J., J. Saric and P. Bork (2006). "Literature mining for the biologist: from information retrieval to biological discovery." Nature reviews genetics **7**(2): 119.

Jenssen, H., P. Hamill and R. E. Hancock (2006). "Peptide antimicrobial agents." Clinical microbiology reviews **19**(3): 491-511.

Ji, X., P.-P. Yao, L.-Y. Zhang, Y. Li, F. Xu, L.-L. Mei, S.-R. Zhu, Y.-J. Zhang, H.-P. Zhu and S. van der Veen (2017). "Capsule switching of Neisseria meningitidis sequence type 7 serogroup A to serogroup X." Journal of Infection **75**(6): 521-531.

John, C. M., N. J. Phillips, D. C. Stein and G. A. Jarvis (2017). "Innate immune response to lipooligosaccharide: pivotal regulator of the pathobiology of invasive Neisseria meningitidis infections." Pathogens and disease **75**(3): ftx030.

Join-Lambert, O., P. C. Morand, E. Carbonnelle, M. Coureuil, E. Bille, S. Bourdoulous and X. Nassif (2010). "Mechanisms of meningeal invasion by a bacterial extracellular pathogen, the example of Neisseria meningitidis." Progress in neurobiology **91**(2): 130-139.

Juba, M. L., D. K. Porter, E. H. Williams, C. A. Rodriguez, S. M. Barksdale and B. M. Bishop (2015). "Helical cationic antimicrobial peptide length and its impact on membrane disruption." Biochimica et Biophysica Acta (BBA)-Biomembranes **1848**(5): 1081-1091.

Juretić, D., A. Tossi, N. Kamech, N. Ilić, V. Bojović, M. Novković, J. Simunić, D. Petrov, B. Lučić and M. Miljak (2013). From Data Collecting to Web servers for Automatic Design of Peptide Antibiotics. Bioinformatics and biological physics: proceedings of the scientific meeting, Hrvatska akademija znanosti i umjetnosti.

Kamble, A. and R. Khairkar "Basics of Bioinformatics in Biological Research." International Journal of Applied Sciences and Biotechnology **4**(4): 425-429.

Kang, S.-J., D.-H. Kim, T. Mishig-Ochir and B.-J. Lee (2012). "Antimicrobial peptides: their physicochemical properties and therapeutic application." Archives of pharmacal research **35**(3): 409-413.

Kauffman, C. and G. Karypis (2010). "Ligand-binding residue prediction." Introduction to Protein Structure Prediction: Methods and Algorithms: 343-368.

Kleine, T. O., P. Zwerenz, P. Zöfel and K. Shiratori (2003). "New and old diagnostic markers of meningitis in cerebrospinal fluid (CSF)." Brain research bulletin **61**(3): 287-297.

Knowles, T. J., A. Scott-Tucker, M. Overduin and I. R. Henderson (2009). "Membrane protein architects: the role of the BAM complex in outer membrane protein assembly." Nature Reviews Microbiology **7**(3): 206.

Kolappan, S., M. Coureuil, X. Yu, X. Nassif, E. H. Egelman and L. Craig (2016). "Structure of the Neisseria meningitidis Type IV pilus." Nature communications **7**: 13015.

Kozlowski, L. P. (2016). "IPC–isoelectric point calculator." Biology direct **11**(1): 55.

Krallinger, M. and A. Valencia (2005). "Text-mining and information-retrieval services for molecular biology." Genome biology **6**(7): 224.

Krause, A., R. Sillard, B. Kleemeier, E. Klüver, E. Maronde, J. R. Conejo-García, W. G. Forssmann, P. Schulz-Knappe, M. C. Nehls and F. Wattler (2003). "Isolation and

biochemical characterization of LEAP-2, a novel blood peptide expressed in the liver." Protein Science **12**(1): 143-152.

Kulagina, N. V., K. M. Shaffer, F. S. Ligler and C. R. Taitt (2007). "Antimicrobial peptides as new recognition molecules for screening challenging species." Sensors and Actuators B: Chemical **121**(1): 150-157.

Kumar, P., J. N. Kizhakkedathu and S. K. Straus (2018). "Antimicrobial Peptides: Diversity, Mechanism of Action and Strategies to Improve the Activity and Biocompatibility In Vivo." Biomolecules **8**(1): 4.

LaForce, F. M., N. Ravenscroft, M. Djingarey and S. Viviani (2009). "Epidemic meningitis due to Group A Neisseria meningitidis in the African meningitis belt: a persistent problem with an imminent solution." Vaccine **27**: B13-B19.

Lai, Y. and R. L. Gallo (2009). "AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense." Trends in immunology **30**(3): 131-141.

Lansac, N., F. Picard, C. Menard, M. Boissinot, M. Ouellette, P. Roy and M. Bergeron (2000). "Novel genus-specific PCR-based assays for rapid identification of Neisseria species and Neisseria meningitidis." European Journal of Clinical Microbiology and Infectious Diseases **19**(6): 443-451.

Larkin, M. A., G. Blackshields, N. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm and R. Lopez (2007). "Clustal W and Clustal X version 2.0." bioinformatics **23**(21): 2947-2948.

Lata, S., B. Sharma and G. Raghava (2007). "Analysis and prediction of antibacterial peptides." BMC bioinformatics **8**(1): 263.

Lee, E. Y., M. W. Lee, B. M. Fulan, A. L. Ferguson and G. C. Wong (2017). "What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning?" Interface focus **7**(6): 20160153.

Leo, J. C., I. Grin and D. Linke (2012). "Type V secretion: mechanism (s) of autotransport through the bacterial outer membrane." Phil. Trans. R. Soc. B **367**(1592): 1088-1101.

Lever, A. and I. Mackenzie (2007). "Sepsis: definition, epidemiology, and diagnosis." Bmj **335**(7625): 879-883.

Li, J., J.-J. Koh, S. Liu, R. Lakshminarayanan, C. S. Verma and R. W. Beuerman (2017). "Membrane active antimicrobial peptides: translating mechanistic insights to design." Frontiers in neuroscience **11**: 73.

Lin, T.-Y. and D. B. Weibel (2016). "Organization and function of anionic phospholipids in bacteria." Applied microbiology and biotechnology **100**(10): 4255-4267.

Lin, X. and X. Zhang (2016). Prediction and analysis of hot region in protein-protein interactions. Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on, IEEE.

Lise, S., C. Archambeau, M. Pontil and D. T. Jones (2009). "Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods." BMC bioinformatics **10**(1): 365.

Liu, S., L. Fan, J. Sun, X. Lao and H. Zheng (2017). "Computational resources and tools for antimicrobial peptides." Journal of Peptide Science **23**(1): 4-12.

Lo, H., C. M. Tang and R. M. Exley (2009). "Mechanisms of avoidance of host immunity by Neisseria meningitidis and its effect on vaccine development." The Lancet infectious diseases **9**(7): 418-427.

London, N., D. Movshovitz-Attias and O. Schueler-Furman (2010). "The structural basis of peptide-protein binding strategies." Structure **18**(2): 188-199.

Lundbo, L. F., H. T. Sørensen, L. N. Clausen, M. V. Hollegaard, D. M. Hougaard, H. B. Konradsen, Z. B. Harboe, M. Nørgaard and T. Benfield (2015). Mannose-binding lectin gene, MBL2, polymorphisms do not increase susceptibility to invasive meningococcal disease in a population of Danish children. Open forum infectious diseases, Oxford University Press.

Luscombe, N. M., D. Greenbaum and M. Gerstein (2001). "What is bioinformatics? An introduction and overview." Yearbook of medical informatics **10**(01): 83-100.

Maharjan, S., M. Saleem, I. M. Feavers, J. X. Wheeler, R. Care and J. P. Derrick (2016). "Dissection of the function of the RmpM periplasmic protein from Neisseria meningitidis." Microbiology **162**(2): 364-375.

Mahlapuu, M., J. Håkansson, L. Ringstad and C. Björn (2016). "Antimicrobial peptides: an emerging category of therapeutic agents." Frontiers in cellular and infection microbiology **6**: 194.

Mahomed, H. and N. Cameron (2006). "Meningococcal disease in Cape Town." Southern African Journal of Epidemiology and Infection **21**(1): 5-8.

Malleshappa Gowder, S., J. Chatterjee, T. Chaudhuri and K. Paul (2014). "Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins." The Scientific World Journal **2014**.

Manchanda, V., S. Gupta and P. Bhalla (2006). "Meningococcal disease: history, epidemiology, pathogenesis, clinical manifestations, diagnosis, antimicrobial susceptibility and prevention." Indian journal of medical microbiology **24**(1): 7.

Mannoor, M. S., S. Zhang, A. J. Link and M. C. McAlpine (2010). "Electrical detection of pathogenic bacteria via immobilized antimicrobial peptides." Proceedings of the National Academy of Sciences **107**(45): 19207-19212.

Matsuzaki, K. (2009). "Control of cell selectivity of antimicrobial peptides." Biochimica et Biophysica Acta (BBA)-Biomembranes **1788**(8): 1687-1692.

McGuffin, L. (2008). Protein fold recognition and threading. Computational Structural Biology: Methods and Applications, World Scientific**:** 37-60.

Meiring, S., G. Hussey, P. Jeena, S. Parker and A. von Gottberg (2017). "Recommendations for the use of meningococcal vaccines in South Africa." Southern African Journal of Infectious Diseases **32**(3): 82-86.

Merz, A. J. and M. So (2000). "Interactions of pathogenic neisseriae with epithelial cell membranes." Annual review of cell and developmental biology **16**(1): 423-457.

Mihajlovic, M. and T. Lazaridis (2010). "Antimicrobial peptides bind more strongly to membrane pores." Biochimica et Biophysica Acta (BBA)-Biomembranes **1798**(8): 1494-1502.

Mihajlovic, M. and T. Lazaridis (2010). "Antimicrobial peptides in toroidal and cylindrical pores." Biochimica et Biophysica Acta (BBA)-Biomembranes **1798**(8): 1485-1493.

Mihajlovic, M. and T. Lazaridis (2012). "Charge distribution and imperfect amphipathicity affect pore formation by antimicrobial peptides." Biochimica et Biophysica Acta (BBA)-Biomembranes **1818**(5): 1274-1283.

Mills, G., H. Lala, M. Oehley, A. Craig, K. Barratt, D. Hood, C. Thornley, A. Nesdale, N. Manikkam and P. Reeve (2006). "Elevated procalcitonin as a diagnostic marker in

meningococcal disease." European Journal of Clinical Microbiology and Infectious Diseases **25**(8): 501-509.

Mojsoska, B. and H. Jenssen (2015). "Peptides and peptidomimetics for antimicrobial drug design." Pharmaceuticals **8**(3): 366-415.

Moore, J. E. (2018). "Meningococcal Disease Section 3: Diagnosis and Management: MeningoNI Forum (see page 87 (2) 83 for full list of authors)." The Ulster medical journal **87**(2): 94.

Moreira, I. S., P. A. Fernandes and M. J. Ramos (2007). "Computational alanine scanning mutagenesis—an improved methodological approach." Journal of Computational Chemistry **28**(3): 644-654.

Moreira, I. S., P. A. Fernandes and M. J. Ramos (2007). "Hot spots—A review of the protein–protein interface determinant amino-acid residues." Proteins: Structure, Function, and Bioinformatics **68**(4): 803-812.

Morris, G. M. and M. Lim-Wilby (2008). Molecular docking. Molecular modeling of proteins, Springer**:** 365-382.

Mulia, S., D. Mishra and T. Jena (2012). "Profile HMM based multiple sequence alignment for DNA sequences." Procedia engineering **38**: 1783-1787.

Mulvenna, J. P., C. Wang and D. J. Craik (2006). "CyBase: a database of cyclic protein sequence and structure." Nucleic acids research **34**(suppl_1): D192-D194.

Nadel, S. and N. Ninis (2018). "Invasive Meningococcal Disease in the Vaccine Era." Frontiers in Pediatrics **6**.

Natale, P., T. Brüser and A. J. Driessen (2008). "Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms." Biochimica et Biophysica Acta (BBA)-Biomembranes **1778**(9): 1735-1756.

Nguyen, L. T., E. F. Haney and H. J. Vogel (2011). "The expanding scope of antimicrobial peptide structures and their modes of action." Trends in biotechnology **29**(9): 464-472.

Nicolas, G. G. (2011). "Detection of putative new mutacins by bioinformatic analysis using available web tools." BioData mining **4**(1): 22.

Nicolau Jr, D. V., E. Paszek, F. Fulga and D. V. Nicolau (2014). "Mapping hydrophobicity on the protein molecular surface at atom-level resolution." PloS one **9**(12): e114042.

Noinaj, N., J. C. Gumbart and S. K. Buchanan (2017). "The β-barrel assembly machinery in motion." Nature Reviews Microbiology **15**(4): 197.

Ofran, Y. (2009). "Prediction of protein interaction sites." Computational Protein-Protein Interactions: 167-184.

Ofran, Y. and B. Rost (2007). "Protein–protein interaction hotspots carved into sequences." PLoS computational biology **3**(7): e119.

Oldfield, N. J., S. Matar, F. A. Bidmos, M. Alamro, K. R. Neal, D. P. Turner, C. D. Bayliss and D. A. Ala'Aldeen (2013). "Prevalence and phase variable expression status of two autotransporters, NalP and MspA, in carriage and disease isolates of Neisseria meningitidis." PloS one **8**(7): e69746.

Oordt-Speets, A. M., R. Bolijn, R. C. van Hoorn, A. Bhavsar and M. H. Kyaw (2018). "Global etiology of bacterial meningitis: A systematic review and meta-analysis." PloS one **13**(6): e0198772.

Organization, W. H. (2011). "Laboratory methods for the diagnosis of meningitis caused by Neisseria meningitidis, Streptococcus pneumoniae and Haemophilus influenzae. WHO manual." Laboratory methods for the diagnosis of meningitis caused by Neisseria meningitidis, Streptococcus pneumoniae and Haemophilus influenzae. WHO manual.(Ed. 2).

Osorio, D., P. Rondón-Villarrea and R. Torres (2015). "Peptides: a package for data mining of antimicrobial peptides." R Journal **7**(1).

Ouzounis, C. A. (2012). "Rise and demise of bioinformatics? Promise and progress." PLoS computational biology **8**(4): e1002487.

Pace, D. and A. J. Pollard (2012). "Meningococcal disease: clinical presentation and sequelae." Vaccine **30**: B3-B9.

Patel, R., S. M. Smith and C. Robinson (2014). "Protein transport by the bacterial Tat pathway." Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1843**(8): 1620-1628.

Patel, S. and N. Akhtar (2017). "Antimicrobial peptides (AMPs): The quintessential 'offense and defense'molecules are more than antimicrobials." Biomedicine & Pharmacotherapy **95**: 1276-1283.

Peak, I. R., A. Chen, F. E.-C. Jen, C. Jennings, B. L. Schulz, N. J. Saunders, A. Khan, H. S. Seifert and M. P. Jennings (2016). "Neisseria meningitidis lacking the major porins PorA and PorB is viable and modulates apoptosis and the oxidative burst of neutrophils." Journal of proteome research **15**(8): 2356-2365.

Pelton, S. I. (2016). "The global evolution of meningococcal epidemiology following the introduction of meningococcal vaccines." Journal of Adolescent Health **59**(2): S3-S11.

Pikora, M. and A. Gieldon (2015). "RASMOL AB-new functionalities in the program for structure analysis." Acta Biochimica Polonica **62**(3).

Pirtskhalava, M., A. Gabrielian, P. Cruz, H. L. Griggs, R. B. Squires, D. E. Hurt, M. Grigolava, M. Chubinidze, G. Gogoladze and B. Vishnepolsky (2015). "DBAASP v. 2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides." Nucleic acids research **44**(D1): D1104-D1112.

Pizza, M. and R. Rappuoli (2015). "Neisseria meningitidis: pathogenesis and immunity." Current opinion in microbiology **23**: 68-72.

Pollard, A. J. and A. Finn (2012). 125 - Neisseria meningitidis A2 - Long, Sarah S. Principles and Practice of Pediatric Infectious Diseases (Fourth Edition). London, Content Repository Only!**:** 730-741.e737.

Polyanovsky, V. O., M. A. Roytberg and V. G. Tumanyan (2011). "Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences." Algorithms for molecular biology **6**(1): 25.

Popovic, T., G. W. Ajello and R. R. Facklam (1998). "Laboratory methods for the diagnosis of meningitis caused by Neisseria meningitidis, Streptococcus pneumoniae, and Haemophilus influenzae."

Porto, W., A. Pires and O. Franco (2017). "Computational tools for exploring sequence databases as a resource for antimicrobial peptides." Biotechnology advances **35**(3): 337-349.

Pushpanathan, M., P. Gunasekaran and J. Rajendhran (2013). "Antimicrobial peptides: versatile biological properties." International journal of peptides **2013**.

Qiao, Y., Y. Xiong, H. Gao, X. Zhu and P. Chen (2018). "Protein-protein interface hot spots prediction based on a hybrid feature selection strategy." BMC bioinformatics **19**(1): 14.

Ramos, R. M. and I. S. Moreira (2013). "Computational Alanine Scanning Mutagenesis · An Improved Methodological Approach for Protein–DNA Complexes." Journal of chemical theory and computation **9**(9): 4243-4256.

Raza, K. (2012). "Application of data mining in bioinformatics." arXiv preprint arXiv:1205.1125.

Ribet, D. and P. Cossart (2015). "How bacterial pathogens colonize their hosts and invade deeper tissues." Microbes and Infection **17**(3): 173-183.

Roche, D. B., D. A. Brackenridge and L. J. McGuffin (2015). "Proteins and their interacting partners: An introduction to protein–ligand binding site prediction methods." International journal of molecular sciences **16**(12): 29829-29842.

Rose, A. M., S. Gerstl, A. E.-H. Mahamane, F. Sidikou, S. Djibo, L. Bonte, D. A. Caugant, P. J. Guerin and S. Chanteau (2009). "Field evaluation of two rapid diagnostic tests for Neisseria meningitidis serogroup A during the 2006 outbreak in Niger." PLoS One **4**(10): e7326.

Rose, A. M., J. E. Mueller, S. Gerstl, B.-M. Njanpop-Lafourcade, A.-L. Page, P. Nicolas, R. O. Traoré, D. A. Caugant and P. J. Guerin (2010). "Meningitis dipstick rapid test: evaluating diagnostic performance during an urban Neisseria meningitidis serogroup A outbreak, Burkina Faso, 2007." PLoS One **5**(6): e11086.

Rosenstein, N. E., B. A. Perkins, D. S. Stephens, L. Lefkowitz, M. L. Cartter, R. Danila, P. Cieslak, K. A. Shutt, T. Popovic and A. Schuchat (1999). "The changing epidemiology of meningococcal disease in the United States, 1992–1996." Journal of Infectious Diseases **180**(6): 1894-1901.

Rosenstein, N. E., B. A. Perkins, D. S. Stephens, T. Popovic and J. M. Hughes (2001). "Meningococcal disease." New England journal of medicine **344**(18): 1378-1388.

Rouphael, N. G. and D. S. Stephens (2012). "Neisseria meningitidis: biology, microbiology, and epidemiology." Neisseria meningitidis: advanced methods and protocols: 1-20.

Roy, A., A. Kucukural and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nature protocols **5**(4): 725.

Sa E Cunha, C., N. J. Griffiths, I. Murillo and M. Virji (2009). "Neisseria meningitidis Opc invasin binds to the cytoskeletal protein α-actinin." Cellular microbiology **11**(3): 389-405.

Sadarangani, M. and A. J. Pollard (2010). "Serogroup B meningococcal vaccines—an unfinished story." The Lancet infectious diseases **10**(2): 112-124.

Sadarangani, M., A. J. Pollard and S. D. Gray-Owen (2011). "Opa proteins and CEACAMs: pathways of immune engagement for pathogenic Neisseria." FEMS microbiology reviews **35**(3): 498-514.

Salmaso, V. and S. Moro (2018). "Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview." Frontiers in pharmacology **9**.

Scarselli, M., D. Serruto, P. Montanari, B. Capecchi, J. Adu-Bobie, D. Veggi, R. Rappuoli, M. Pizza and B. Aricò (2006). "Neisseria meningitidis NhhA is a multifunctional trimeric autotransporter adhesin." Molecular microbiology **61**(3): 631-644.

Schauperl, M., M. Podewitz, B. J. Waldner and K. R. Liedl (2016). "Enthalpic and entropic contributions to hydrophobicity." Journal of chemical theory and computation **12**(9): 4600-4610.

Schibli, D. J., H. N. Hunter, V. Aseyev, T. D. Starner, J. M. Wiencek, P. B. McCray, B. F. Tack and H. J. Vogel (2002). "The solution structures of the human β-defensins lead to a

better understanding of the potent bactericidal activity of HBD3 against Staphylococcus aureus." Journal of Biological Chemistry **277**(10): 8279-8289.

Schluesener, H. J., Y. Su, A. Ebrahimi and D. Pouladsaz (2012). "Antimicrobial peptides in the brain: neuropeptides and amyloid." Front Biosci (Schol Ed) **4**: 1375-1380.

Schmidtchen, A., M. Pasupuleti and M. Malmsten (2014). "Effect of hydrophobic modifications in antimicrobial peptides." Advances in colloid and interface science **205**: 265-274.

Schneidman-Duhovny, D., Y. Inbar, R. Nussinov and H. J. Wolfson (2005). "PatchDock and SymmDock: servers for rigid and symmetric docking." Nucleic acids research **33**(suppl_2): W363-W367.

Schwede, T. (2013). "Protein modeling: what happened to the "protein structure gap"?" Structure **21**(9): 1531-1540.

Schwede, T., A. Sali, N. Eswar and M. Peitsch (2008). Protein structure modeling. Computational Structural Biology: Methods and Applications, World Scientific**:** 3-35.

Seib, K. L., M. Scarselli, M. Comanducci, D. Toneatto and V. Masignani (2015). "Neisseria meningitidis factor H-binding protein fHbp: a key virulence factor and vaccine antigen." Expert review of vaccines **14**(6): 841-859.

Selkrig, J., K. Mosbahi, C. T. Webb, M. J. Belousoff, A. J. Perry, T. J. Wells, F. Morris, D. L. Leyton, M. Totsika and M.-D. Phan (2012). "Discovery of an archetypal protein transport system in bacterial outer membranes." Nature Structural and Molecular Biology **19**(5): 506.

Serruto, D., T. Spadafina, L. Ciucchi, L. A. Lewis, S. Ram, M. Tontini, L. Santini, A. Biolchi, K. L. Seib and M. M. Giuliani (2010). "Neisseria meningitidis GNA2132, a heparin-

binding protein that induces protective immunity in humans." Proceedings of the National Academy of Sciences **107**(8): 3770-3775.

Seshadri Sundararajan, V., M. N. Gabere, A. Pretorius, S. Adam, A. Christoffels, M. Lehväslaiho, J. A. Archer and V. B. Bajic (2011). "DAMPD: a manually curated antimicrobial peptide database." Nucleic acids research **40**(D1): D1108-D1112.

Shai, Y. (2002). "Mode of action of membrane active antimicrobial peptides." Peptide Science **66**(4): 236-248.

Shams, F., N. J. Oldfield, S. K. Lai, S. A. Tunio, K. G. Wooldridge and D. P. Turner (2016). "Fructose-1, 6-bisphosphate aldolase of Neisseria meningitidis binds human plasminogen via its C-terminal lysine residue." MicrobiologyOpen **5**(2): 340-350.

Simonis, A. and A. Schubert-Unkmeir (2016). "Interactions of meningococcal virulence factors with endothelial cells at the human blood–cerebrospinal fluid barrier and their role in pathogenicity." FEBS letters **590**(21): 3854-3867.

Simpson, M. S. and D. Demner-Fushman (2012). Biomedical text mining: a survey of recent progress. Mining text data, Springer**:** 465-517.

Sjölinder, H., J. Eriksson, L. Maudsdotter, H. Aro and A.-B. Jonsson (2008). "Meningococcal outer membrane protein NhhA is essential for colonization and disease by preventing phagocytosis and complement attack." Infection and immunity **76**(11): 5412-5420.

Sjölinder, M., G. Altenbacher, M. Hagner, W. Sun, S. Schedin-Weiss and H. Sjölinder (2012). "Meningococcal outer membrane protein NhhA triggers apoptosis in macrophages." PloS one **7**(1): e29586.

Skewes-Cox, P., T. J. Sharpton, K. S. Pollard and J. L. DeRisi (2014). "Profile hidden Markov models for the detection of viruses within metagenomic sequence data." PLoS One **9**(8): e105067.

Sousa, S. A., J. H. Leitão, R. C. Martins, J. M. Sanches, J. S. Suri and A. Giorgetti (2016). "Bioinformatics applications in life sciences and technologies." BioMed research international **2016**.

Stephens, D. S. (2007). "Conquering the meningococcus." FEMS microbiology reviews **31**(1): 3-14.

Stephens, D. S. (2009). "Biology and pathogenesis of the evolutionarily successful, obligate human bacterium Neisseria meningitidis." Vaccine **27**: B71-B77.

Stephens, D. S., B. Greenwood and P. Brandtzaeg (2007). "Epidemic meningitis, meningococcaemia, and Neisseria meningitidis." The Lancet **369**(9580): 2196-2210.

Swartley, J., A. Marfin, S. Edupuganti, L.-J. Liu, P. Cieslak, B. Perkins, J. Wenger and D. Stephens (1997). "Capsule switching of Neisseria meningitidis." Proceedings of the National Academy of Sciences **94**(1): 271-276.

Takada, S., S. Fujiwara, T. Inoue, Y. Kataoka, Y. Hadano, K. Matsumoto, K. Morino and T. Shimizu (2016). "Meningococcemia in adults: a review of the literature." Internal Medicine **55**(6): 567-572.

Talà, A., C. Progida, M. De Stefano, L. Cogli, M. R. Spinosa, C. Bucci and P. Alifano (2008). "The HrpB–HrpA two-partner secretion system is essential for intracellular survival of Neisseria meningitidis." Cellular microbiology **10**(12): 2461-2482.

Tang, Y., X. Zeng and J. Liang (2010). "Surface plasmon resonance: an introduction to a surface spectroscopy technique." Journal of chemical education **87**(7): 742-746.

Tatusova, T., M. DiCuccio, A. Badretdin, V. Chetvernin, S. Ciufo and W. Li (2013). "The NCBI handbook." National Center for Biotechnology Information.

Terrade, A., J.-M. Collard, F. Nato and M.-K. Taha (2013). "Laboratory evaluation of a rapid diagnostic test for Neisseria meningitidis serogroup A." Transactions of the Royal Society of Tropical Medicine and Hygiene **107**(7): 460-461.

Terrin, G., A. Passariello, F. Manguso, G. Salvia, L. Rapacciuolo, F. Messina, F. Raimondi and R. B. Canani (2011). "Serum calprotectin: an antimicrobial peptide as a new marker for the diagnosis of sepsis in very low birth weight newborns." Clinical and Developmental Immunology **2011**.

Thomas, S., S. Karnik, R. S. Barai, V. K. Jayaraman and S. Idicula-Thomas (2009). "CAMP: a useful resource for research on antimicrobial peptides." Nucleic acids research **38**(suppl_1): D774-D780.

Thompson, M. J., N. Ninis, R. Perera, R. Mayon-White, C. Phillips, L. Bailey, A. Harnden, D. Mant and M. Levin (2006). "Clinical recognition of meningococcal disease in children and adolescents." The lancet **367**(9508): 397-403.

Tincho, M., M. Gabere and A. Pretorius (2016). "In silico identification and molecular validation of putative antimicrobial peptides for HIV therapy." Journal of AIDS and Clinical Research **7**(9).

Tincho, M. B. (2016). "In-silico optimization and molecular validation of putative anti-HIV antimicrobial peptides for therapeutic purpose."

Tokuda, H. (2009). "Biogenesis of outer membranes in Gram-negative bacteria." Bioscience, biotechnology, and biochemistry **73**(3): 465-473.

Tommassen, J. and J. Arenas (2017). "Biological Functions of the Secretome of Neisseria meningitidis." Frontiers in cellular and infection microbiology **7**: 256.

Toneatto, D., M. Pizza, V. Masignani and R. Rappuoli (2017). "Emerging experience with meningococcal serogroup B protein vaccines." Expert review of vaccines **16**(5): 433-451.

Torrent, M., D. Andreu, V. M. Nogués and E. Boix (2011). "Connecting peptide physicochemical and antimicrobial properties by a rational prediction model." PloS one **6**(2): e16968.

Tossi, A., L. Sandri and A. Giangaspero (2000). "Amphipathic, α-helical antimicrobial peptides." Peptide Science **55**(1): 4-30.

Travkova, O. G., H. Moehwald and G. Brezesinski (2017). "The interaction of antimicrobial peptides with membranes." Advances in colloid and interface science **247**: 521-532.

Tsang, R. and M.-K. Taha (2016). Diagnosis of meningococcal disease. Handbook of Meningococcal Disease Management, Springer**:** 45-55.

Tuncbag, N., A. Gursoy and O. Keskin (2009). "Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy." Bioinformatics **25**(12): 1513-1520.

Tuncbag, N., G. Kar, O. Keskin, A. Gursoy and R. Nussinov (2009). "A survey of available tools and web servers for analysis of protein–protein interactions and interfaces." Briefings in bioinformatics **10**(3): 217-232.

Tunio, S. A., N. J. Oldfield, D. A. Ala'Aldeen, K. G. Wooldridge and D. P. Turner (2010). "The role of glyceraldehyde 3-phosphate dehydrogenase (GapA-1) in Neisseria meningitidis adherence to human cells." BMC microbiology **10**(1): 280.

Tunio, S. A., N. J. Oldfield, A. Berry, D. A. Ala'Aldeen, K. G. Wooldridge and D. P. Turner (2010). "The moonlighting protein fructose-1, 6-bisphosphate aldolase of Neisseria meningitidis: surface localization and role in host cell adhesion." Molecular microbiology **76**(3): 605-615.

Tzeng, Y.-L. and D. S. Stephens (2000). "Epidemiology and pathogenesis of Neisseria meningitidis." Microbes and infection **2**(6): 687-700.

Tzeng, Y.-L. and D. S. Stephens (2015). "Antimicrobial peptide resistance in Neisseria meningitidis." Biochimica et Biophysica Acta (BBA)-Biomembranes **1848**(11): 3026-3031.

Uadiale, K., A. Bestman, C. Kamau, D. A. Caugant and J. Greig (2016). "Evaluation of Pastorex meningitis kit performance for the rapid identification of Neisseria meningitidis serogroup C in Nigeria." Transactions of The Royal Society of Tropical Medicine and Hygiene **110**(7): 381-385.

Uberos, J., M. Molina-Oya, S. Martinez-Serrano and L. Fernández-López (2015). "Surface adhesion and host response as pathogenicity factors of Neisseria meningitidis." World Journal of Clinical Infectious Diseases **5**(2): 37-43.

Urfer, M., J. Bogdanovic, F. L. Monte, K. Moehle, K. Zerbe, U. Omasits, C. H. Ahrens, G. Pessi, L. Eberl and J. A. Robinson (2016). "A peptidomimetic antibiotic targets outer membrane proteins and disrupts selectively the outer membrane in Escherichia coli." Journal of Biological Chemistry **291**(4): 1921-1932.

Uria, M. J., Q. Zhang, Y. Li, A. Chan, R. M. Exley, B. Gollan, H. Chan, I. Feavers, A. Yarwood and R. Abad (2008). "A generic mechanism in Neisseria meningitidis for enhanced resistance against bactericidal antibodies." Journal of Experimental Medicine **205**(6): 1423-1434.

Van Landeghem, S., J. Björne, T. Abeel, B. De Baets, T. Salakoski and Y. Van de Peer (2012). Semantically linking molecular entities in literature through entity relationships. BMC bioinformatics, BioMed Central.

van Ulsen, P. (2011). Protein folding in bacterial adhesion: secretion and folding of classical monomeric autotransporters. Bacterial Adhesion, Springer**:** 125-142.

Vanhee, P., A. M. van der Sloot, E. Verschueren, L. Serrano, F. Rousseau and J. Schymkowitz (2011). "Computational design of peptide receptors." Trends in biotechnology **29**(5): 231-239.

Vanhoye, D., F. Bruston, S. El Amri, A. Ladram, M. Amiche and P. Nicolas (2004). "Membrane association, electrostatic sequestration, and cytotoxicity of Gly-Leu-rich peptide orthologs with differing functions." Biochemistry **43**(26): 8391-8409.

Varkey, J. and R. Nagaraj (2005). "Antibacterial activity of human neutrophil defensin HNP-1 analogs without cysteines." Antimicrobial agents and chemotherapy **49**(11): 4561-4566.

Vázquez, J., M. Taha, J. Findlow, S. Gupta and R. Borrow (2016). "Global Meningococcal Initiative: guidelines for diagnosis and confirmation of invasive meningococcal disease." Epidemiology & Infection **144**(14): 3052-3057.

Veltri, D., U. Kamath and A. Shehu (2018). "Deep learning improves antimicrobial peptide recognition." Bioinformatics **1**: 8.

Veltri, D. P. (2015). A Computational and Statistical Framework for Screening Novel Antimicrobial Peptides.

Vidal, J. E. and D. R. Boulware (2015). "Lateral flow assay for cryptococcal antigen: an important advance to improve the continuum of HIV care and reduce cryptococcal

meningitis-related mortality." Revista do Instituto de Medicina Tropical de São Paulo **57**: 38-45.

Vijayan, A. L., S. Ravindran, R. Saikant, S. Lakshmi and R. Kartik (2017). "Procalcitonin: a promising diagnostic marker for sepsis and antibiotic therapy." Journal of intensive care **5**(1): 51.

Virji, M. (2009). "Pathogenic neisseriae: surface modulation, pathogenesis and infection control." Nature Reviews Microbiology **7**(4): 274.

Vishnepolsky, B., A. Gabrielian, A. Rosenthal, D. E. Hurt, M. Tartakovsky, G. Managadze, M. Grigolava, G. I. Makhatadze and M. Pirtskhalava (2018). "Predictive Model of Linear Antimicrobial Peptides Active against Gram-Negative Bacteria." Journal of chemical information and modeling **58**(5): 1141-1151.

Von Gottberg, A., M. Du Plessis, C. Cohen, E. Prentice, S. Schrag, L. de Gouveia, G. Coulson, G. de Jong, K. Klugman, R. Group for Enteric and M. D. S. i. S. Africa (2008). "Emergence of endemic serogroup W135 meningococcal disease associated with a high mortality rate in South Africa." Clinical Infectious Diseases **46**(3): 377-386.

Vos, T., A. A. Abajobir, K. H. Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, R. S. Abdulkader, A. M. Abdulle, T. A. Abebo and S. F. Abera (2017). "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016." The Lancet **390**(10100): 1211-1259.

Waghu, F. H., R. S. Barai, P. Gurung and S. Idicula-Thomas (2015). "CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides." Nucleic acids research **44**(D1): D1094-D1097.

Waghu, F. H., R. S. Barai and S. Idicula-Thomas (2016). "Leveraging family-specific signatures for AMP discovery and high-throughput annotation." Scientific reports **6**: 24684.

Waghu, F. H., L. Gopi, R. S. Barai, P. Ramteke, B. Nizami and S. Idicula-Thomas (2014). "CAMP: Collection of sequences and structures of antimicrobial peptides." Nucleic acids research **42**(D1): D1154-D1158.

Wang, C. K., Q. Kaas, L. Chiche and D. J. Craik (2007). "CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering." Nucleic acids research **36**(suppl_1): D206-D210.

Wang, G. (2014). "Human antimicrobial peptides and proteins." Pharmaceuticals **7**(5): 545-594.

Wang, G. (2015). "Improved methods for classification, prediction, and design of antimicrobial peptides." Computational Peptidology, Springer**:** 43-66.

Wang, G. (2017). "Antimicrobial peptides: discovery, design and novel therapeutic strategies". Cabi

Wang, G., X. Li and Z. Wang (2009). "APD2: The updated antimicrobial peptide database and its application in peptide design." Nucleic Acids Research **37**(SUPPL. 1).

Wang, G., X. Li and Z. Wang (2015). "APD3: the antimicrobial peptide database as a tool for research and education." Nucleic acids research **44**(D1): D1087-D1093.

Wang, G., X. Li and Z. Wang (2016). "APD3: the antimicrobial peptide database as a tool for research and education." Nucleic acids research **44**(D1): D1087-1093.

Wang, G., Y. Xia, J. Cui, Z. Gu, Y. Song, Y. Chen, H. Chen, H. Zhang and W. Chen (2014). "The Roles of Moonlighting Proteins in Bacteria." Current issues in molecular biology **16**: 15.

Wang, H., C. Liu and L. Deng (2018). "Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting." Scientific reports **8**(1): 14285.

Wang, L., Z.-P. Liu, X.-S. Zhang and L. Chen (2012). "Prediction of hot spots in protein interfaces using a random forest model with hybrid features." Protein Engineering, Design & Selection **25**(3): 119-126.

Wang, P., L. Hu, G. Liu, N. Jiang, X. Chen, J. Xu, W. Zheng, L. Li, M. Tan and Z. Chen (2011). "Prediction of antimicrobial peptides based on sequence alignment and feature selection methods." PloS one **6**(4): e18476.

Wang, Z. and G. Wang (2004). "APD: the antimicrobial peptide database." Nucleic acids research **32**(suppl_1): D590-D592.

Williams, D. A., T. Kiiza, R. Kwizera, R. Kiggundu, S. Velamakanni, D. B. Meya, J. Rhein and D. R. Boulware (2015). "Evaluation of fingerstick cryptococcal antigen lateral flow assay in HIV-infected persons: a diagnostic accuracy study." Clinical Infectious Diseases **61**(3): 464-467.

Williams, M., M. Tincho, M. Gabere, A. Uys and M. Meyer (2016). "Molecular validation of putative antimicrobial peptides for improved Human Immunodeficiency Virus diagnostics via HIV protein p24." J. AIDS Clin. Res **7**: 571.

Wilson, M. M. and H. D. Bernstein (2016). "Surface-exposed lipoproteins: an emerging secretion phenomenon in Gram-negative bacteria." Trends in microbiology **24**(3): 198-208.

Woo, H.-J. (2008). Calculation of absolute protein–ligand binding constants with the molecular dynamics free energy perturbation method. Molecular Modeling of Proteins, Springer**:** 109-120.

Wu, C. H. and U. Consortium (2016). "UniProt: the universal protein knowledgebase."

Wylie, P. A., D. Stevens, W. Drake, J. Stuart and K. Cartwright (1997). "Epidemiology and clinical management of meningococcal disease in west Gloucestershire: retrospective, population based study." Bmj **315**(7111): 774-779.

Xia, J., Z. Yue, Y. Di, X. Zhu and C.-H. Zheng (2016). "Predicting hot spots in protein interfaces based on protrusion index, pseudo hydrophobicity and electron-ion interaction pseudopotential features." Oncotarget **7**(14): 18065.

Xie, O., A. J. Pollard, J. E. Mueller and G. Norheim (2013). "Emergence of serogroup X meningococcal disease in Africa: need for a vaccine." Vaccine **31**(27): 2852-2861.

Yadav, S., S. K. Pandey, V. K. Singh, Y. Goel, A. Kumar and S. M. Singh (2017). "Molecular docking studies of 3-bromopyruvate and its derivatives to metabolic regulatory enzymes: Implication in designing of novel anticancer therapeutic strategies." PloS one **12**(5): e0176403.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang (2015). "The I-TASSER Suite: protein structure and function prediction." Nature methods **12**(1): 7.

Yang, J. and Y. Zhang (2015). "I-TASSER server: new development for protein structure and function predictions." Nucleic acids research **43**(W1): W174-W181.

Yang, J. and Y. Zhang (2015). "Protein structure and function prediction using I-TASSER." Current protocols in bioinformatics **52**(1): 5.8. 1-5.8. 15.

Yazdankhah, S. P. and D. A. Caugant (2004). "Neisseria meningitidis: an overview of the carriage state." Journal of medical microbiology **53**(9): 821-832.

Yeaman, M. R. and N. Y. Yount (2003). "Mechanisms of antimicrobial peptide action and resistance." Pharmacological reviews **55**(1): 27-55.

Yezli, S., A. M. Assiri, R. F. Alhakeem, A. M. Turkistani and B. Alotaibi (2016). "Meningococcal disease during the Hajj and Umrah mass gatherings." International Journal of Infectious Diseases **47**: 60-64.

Yoon, B.-J. (2009). "Hidden Markov models and their applications in biological sequence analysis." Current genomics **10**(6): 402-415.

Yunta, M. (2016). "Docking and ligand binding affinity: uses and pitfalls." Am J Model Optim **4**: 74-114.

Yuriev, E., M. Agostino and P. A. Ramsland (2011). "Challenges and advances in computational docking: 2009 in review." Journal of Molecular Recognition **24**(2): 149-164.

Yuriev, E., J. Holien and P. A. Ramsland (2015). "Improvements, trends, and new ideas in molecular docking: 2012–2013 in review." Journal of Molecular Recognition **28**(10): 581-604.

Zasloff, M. (2002). "Antimicrobial peptides of multicellular organisms." nature **415**(6870): 389.

Zelezetsky, I., A. Pontillo, L. Puzzi, N. Antcheva, L. Segat, S. Pacor, S. Crovella and A. Tossi (2006). "Evolution of the Primate Cathelicidin Correlation between Structural Variations and Antimicrobial Activity." Journal of Biological Chemistry **281**(29): 19861-19871.

Zhang, L. and R. Gallo (2016). "Antimicrobial peptides." Current biology: CB **26**(1): R14-19.

Zhang, L., A. Rozek and R. E. Hancock (2001). "Interaction of cationic antimicrobial peptides with model membranes." Journal of Biological Chemistry.

Zhang, S.-K., J.-w. Song, F. Gong, S.-B. Li, H.-Y. Chang, H.-M. Xie, H.-W. Gao, Y.-X. Tan and S.-P. Ji (2016). "Design of an α-helical antimicrobial peptide with improved cell-selective and potent anti-biofilm activity." Scientific reports **6**: 27394.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC bioinformatics **9**(1): 40.

Zhou, H.-X. and S. Qin (2007). "Interaction-site prediction for protein complexes: a critical assessment." Bioinformatics **23**(17): 2203-2209.

Zhu, X. and J. C. Mitchell (2011). "KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features." Proteins: Structure, Function, and Bioinformatics **79**(9): 2671-2683.

Zou, D., L. Ma, J. Yu and Z. Zhang (2015). "Biological databases for human research." Genomics, proteomics & bioinformatics **13**(1): 55-63.

APPENDICES

Appendix A

Supplementary Material for Chapter 2

Table A.1: HMMER scores and E-values for AMPs

Single domain	All AMPs	
name	Scores	E-value
YYNN1	33.8	3.50E-06
YYNN2	23.2	0.0033
YYNN3	26	0.00085
YYNN4	24.1	0.0067
YYNN5	30.5	2.90E-05
YYNN6	32.6	2.00E-05
YYNN7	29.9	0.00013
	N.gon	
YYNN8	37.6	1.50E-07
	N.men	
YYNN9	29.1	2.40E-06

Appendix B

Supplementary Material for Chapter 3

Table B.1: NCBI result for NhhA

```
>NP_274028.1 adhesin - nhhA
MNKIYRIIWNSALNAWVVVSELTRNHTKRASATVKTAVLATLLFATVQASANNEEQEEDLYLDPVQRTVA
VLIVNSDKEGTGEKEKVEENSDWAVYFNEKGVLTAREITLKAGDNLKIKQNGTNFTYSLKKDLTDLTSVG
TEKLSFSANGNKVNITSDTKGLNFAKETAGTNGDTTVHLNGIGSTLTDTLLNTGATTNVTNDNVTDDEKK
RAASVKDVLNAGWNIKGVKPGTTASDNVDFVRTYDTVEFLSADTKTTTVNVESKDNGKKTEVKIGAKTSV
IKEKDGKLVTGKDKGENGSSTDEGEGLVTAKEVIDAVNKAGWRMKTTTANGQTGQADKFETVTSGTNVTF
ASGKGTTATVSKDDQGNITVMYDVNVGDALNVNQLQNSGWNLDSKAVAGSSGKVISGNVSPSKGKMDETV
NINAGNNIEITRNGKNIDIATSMTPQFSSVSLGAGADAPTLSVDGDALNVGSKKDNKPVRITNVAPGVKE
GDVTNVAQLKGVAQNLNNRIDNVDGNARAGIAQAIATAGLVQAYLPGKSMMAIGGGTYRGEAGYAIGYSS
ISDGGNWIIKGTASGNSRGHFGASASVGYQW
```

Table B.2: UniProt result for NhhA

```
>sp|Q7DDJ2|NHHA_NEIMB Autotransporter adhesin NhhA
MNKIYRIIWNSALNAWVVVSELTRNHTKRASATVKTAVLATLLFATVQASANNEEQEEDLYLDPVQRTVA
VLIVNSDKEGTGEKEKVEENSDWAVYFNEKGVLTAREITLKAGDNLKIKQNGTNFTYSLKKDLTDLTSVG
TEKLSFSANGNKVNITSDTKGLNFAKETAGTNGDTTVHLNGIGSTLTDTLLNTGATTNVTNDNVTDDEKK
RAASVKDVLNAGWNIKGVKPGTTASDNVDFVRTYDTVEFLSADTKTTTVNVESKDNGKKTEVKIGAKTSV
IKEKDGKLVTGKDKGENGSSTDEGEGLVTAKEVIDAVNKAGWRMKTTTANGQTGQADKFETVTSGTNVTF
```

ASGKGTTATVSKDDQGNITVMYDVNVGDALNVNQLQNSGWNLDSKAVAGSSGKVISGNVSPSKGKMDETV
NINAGNNIEITRNGKNIDIATSMTPQFSSVSLGAGADAPTLSVDGDALNVGSKKDNKPVRITNVAPGVKE
GDVTNVAQLKGVAQNLNNRIDNVDGNARAGIAQAIATAGLVQAYLPGKSMMAIGGGTYRGEAGYAIGYSS
ISDGGNWIIKGTASGNSRGHFGASASVGYQW

Table B.3: NCBI result for PorA

>NP_274441.1 outer membrane protein PorA
MRKKLTALVLSALPLAAVADVSLYGEIKAGVEGRNYQLQLTEAQAANGGASGQVKVTKVT
KAKSRIRTKISDFGSFIGFKGSEDLGDGLKAVWQLEQDVSVAGGGATQWGNRESFIGLAG
EFGTLRAGRVANQFDDASQAIDPWDSNNDVASQLGIFKRHDDMPVSVRYDSPEFSGFSGS
VQFVPIQNSKSAYTPAYYTKNTNNNLTLVPAVVGKPGSDVYYAGLNYKNGGFAGNYAFKY
ARHANVGRNAFELFLIGSGSDQAKGTDPLKNHQVHRLTGGYEEGGLNLALAAQLDLSENG
DKTKNSTTEIAATASYRFGNAVPRISYAHGFDFIERGKKGENTSYDQIIAGVDYDFSKRT
SAIVSGAWLKRNTGIGNYTQINAASVGLRHKF

Table B.4: UniProt result for PorA

>sp|P0DH58|OMPA_NEIMB Major outer membrane protein PorA
MRKKLTALVLSALPLAAVADVSLYGEIKAGVEGRNYQLQLTEAQAANGGASGQVKVTKVT
KAKSRIRTKISDFGSFIGFKGSEDLGDGLKAVWQLEQDVSVAGGGATQWGNRESFIGLAG
EFGTLRAGRVANQFDDASQAIDPWDSNNDVASQLGIFKRHDDMPVSVRYDSPEFSGFSGS
VQFVPIQNSKSAYTPAYYTKNTNNNLTLVPAVVGKPGSDVYYAGLNYKNGGFAGNYAFKY
ARHANVGRNAFELFLIGSGSDQAKGTDPLKNHQVHRLTGGYEEGGLNLALAAQLDLSENG
DKTKNSTTEIAATASYRFGNAVPRISYAHGFDFIERGKKGENTSYDQIIAGVDYDFSKRT
SAIVSGAWLKRNTGIGNYTQINAASVGLRHKF

Table B.5: NCBI result for Opc

>NP_274087.1 class 5 outer membrane protein - Opc
MKKTVFTCAMIALTGTAAAAQELQTANEFTVHTDLSSISSTRAFLKEKHKAAKHISVRAD
IPFDANQGIRLEAGFGRSKKNIINLETDENKLGKTKNVKLPTGVPENRIDLYTGYTYTQT
LSDSLNFRVGAGLGFESSKDSIKTTKHTLHSSRQSWLAKVHADLLSQLGNGWYINPWSEV
KFDLNSRYKLNTGVTNLKKDINQKTNGWGFGLGANIGKKLGESASIEAGPFYKQRTYKES
GEFSVTTKSGDVSLTIPKTSIREYGLRVGIKF

Table B.6: UniProt result for Opc

>tr|Q51227|Q51227_NEIME OpcA protein OS=Neisseria meningitidis
MKKTVFTCAMIALTGTAAAAQELQTANEFTVHTDLSSISSTRAFLKEKHKAAKHIGVRAD
IPFDANQGIRLEAGFGRSKKNIINLETDENKLGKTKNVKLPTGVPENRIDLYTGYTYTQT
LSDSLNFRVGAGLGFESSKDSIKTTKHTLHSSRQSWLAKVHADLLSQLGNGWYINPWSEV
KFDLNSRYKLNTGVTNLKKDINQKTNGWGFGLGANIGKKLGESASIEAGPFYKQRTYKES
GEFSVTTKSGDVSLTIPKTSIREYGLRVGIKF